2014

Pronominal Complex Predicates in Colloquial Persian

Ghazaleh Kazeminejad
University of Kentucky, gh.kazeminejad@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Ghazaleh Kazeminejad, Student

Dr. Gregory Stump, Major Professor

Dr. Gregory Stump, Director of Graduate Studies

PRONOMINAL COMPLEX PREDICATES IN COLLOQUIAL PERSIAN

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Arts in Linguistics
in the College of Arts and Sciences
at the University of Kentucky

By

Ghazaleh Kazeminejad

Lexington, Kentucky

Co-Directors: Dr. Gregory Stump, Professor of Linguistics
and Dr. Andrew Hippisley, Professor of Linguistics

Lexington, Kentucky

2014

ABSTRACT OF THESIS

PRONOMINAL COMPLEX PREDICATES IN COLLOQUIAL PERSIAN

Pronominal complex predicates in colloquial Persian are periphrastic constructions with an idiosyncratic syntactic pattern. They show a peculiar behavior compared to the regular agreement system in Persian, and they are the only construction in Persian which requires the obligatory presence of a pronominal enclitic. This work is an attempt to analyze this construction in order to find its function. For this purpose, a lexical semantic classification of them was proposed, which helped in presenting a new analysis. It was found out that this construction is used to express a particular diathesis in which the topic of the sentence (determined according to Givón's topicality hierarchy) is an indirect participant. I proposed a hybrid dual-layer agreement system which includes a morphosyntactic and a semantic layer. The pronominal enclitic was analyzed as a phrasal affix and agreement marker by reference to Givón's (1976) and Anderson's (2005) arguments. The construction was analyzed to be an instance of the external possessor construction proposed by Haig (2008), which is observed in Iranian languages. The classification of the data clarified the mapping of semantics onto syntax. The proposed analysis could be added to and unified with the current analysis of Persian complex predicates (Bonami and Samvelian, 2009).

KEYWORDS: Indirect Participation, Hybrid Agreement, External Possessor Construction, Iranian Linguistics, Syntax-Semantic Interface

Ghazaleh Kazeminejad

July 24, 2014

PRONOMINAL COMPLEX PREDICATES IN COLLOQUIAL PERSIAN

By

Ghazaleh Kazeminejad

Dr. Gregory Stump
Co-Director of Thesis

Dr. Andrew Hippisley
Co-Director of Thesis

Gregory Stump
Director of Graduate Studies

July 24, 2014

DEDICATION

To my parents,

Fereshteh and Majid.

ACKNOWLEDGEMENTS

This work could not have been carried out without the insights and advices of a number of people. As the chair of my committee, Dr. Gregory Stump provided instructive comments in every stage of this research, directing me to keep in the right track. His discerning recommendations, both in the broader perspective and in the details, helped in developing and enhancing the analysis proposed in this work. In addition, Dr. Andrew Hippisley taught me how to write a thesis and provided invaluable comments on the thesis. I owe a debt of gratitude to Dr. Pollet Samvelian, the third member of my thesis committee, for inspiring me all throughout this work. With her vast knowledge on Persian linguistics, complex predicates, and more generally Iranian linguistics, she provided priceless comments and observations which illuminated the less clear aspects of this work. I would also express my gratitude to Dr. Fabiola Henry for her fabulous suggestions which changed my perspective and led me to new ways. I also wish to thank Dr. Simin Karimi for her helpful comments on my work, and Sedigheh Moradi for having continuous intellectual dialogues about my work.

In addition to the academic support mentioned above, I'm indebted to Mahnaz and Mehdi who supported me emotionally in the absence of my parents, Fereshteh and Majid Kazeminejad, who have always encouraged and supported me to pursue academic studies. I wish to thank my best friends, Mahnaz, Maryam and Neda, who warmed my heart with their love.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

List of Tables vii

List of Figures viii

List of Abbreviations viii

Chapter 1 - Introduction 1

1.1 The issue 2

1.2 Methodology 6

1.3 Outline of the thesis 7

Chapter 2 - Literature Review 9

2.1 Introduction 9

2.2 Diachronic Studies 10

2.3 Verbal Structure 11

2.4 Agreement/Clitic 18

2.5 Subject/Topic 22

Chapter 3 – Theoretical Considerations 25

3.1 Valence Alternation 26

3.1.1 Conversive 28

3.1.2 Anticausative 29

3.2 Complex (-or Periphrastic-) Predicates 30

3.3 Subject Agreement 33

3.3.1 Topichood 35

3.3.2 Topic Agreement 36

3.4 Pronominal Clitics vs. Morphological Affixes 39

3.5 External Possession Construction 42

Chapter 4 – Data Analysis 47

4.1 Introduction 47

4.2 Lexical Semantic Analysis of the Data 48

4.2.1 Possessor Centered Self 48

4.2.1.1 Verbs of Memory 48

4.2.1.2 Verbs of Desire 51

4.2.1.3 Conjecture Verbs 53

4.2.2 Dative Centered Self 54

4.2.2.1 Possessive Dative 54

4.2.2.2 Psych Predicates 58

4.2.2.3 Long Verbs 60

4.2.2.4 Verbs Involving the Body 60

4.2.2.5 Verbs of Disappearance 62

4.2.2.6 Correspond Verbs 62

4.2.2.7 Verbs of Relation 63

4.3 Light-Verb-Based Classification 64

4.3.1 The Copulas: *šodan/budan* 66

4.3.2 Other Light Verbs 69

4.3.2.1 gereftan ('to take') 69

4.3.2.2 umadan ('to come') 72

4.3.2.3 bordan ('to carry') 74

4.3.2.4 zadan ('to hit, to strike') 76

4.4 Summary 77

Chapter 5 - Application of Theory to Persian Data 78

5.1 Diathesis Alternation 78

5.2 Agreement 81

5.3 Pronominal Clitics 83

Chapter 6 – Conclusion 85

6.1 Summary and Conclusion 85

6.2 Recommendations for further study.....*86*

Appendix-1 The PCPs in the corpus: light-verb-based classification 87

Appendix-2 Alphabetical ordering of the PCPs with glosses 89

References 92

Vita.....95

List of Tables

Table 1-1: Personal Endings 2

Table 1-2: Simple past conjugation of *xordan* ‘to eat’ 2

Table 1-3: Simple present conjugation of the compound verb *dust dâštan* ‘to like’ 3

Table 1-4: PCP conjugation of *xoš umadan* ‘to like’ 3

Table 1-5: Object pronominal enclitics 4

Table 2-1: Windfuhr’s classification of PCPs based on Khanlari’s data 12

Table 4-1: verbs of memory 50

Table 4-2: the most frequent light verbs used in PCPs 64

Table 4-3: less frequent light verbs used in some limited PCPs 65

List of Figures

Figure 3-1: The function of the Old Persian Genitive..... 45

List of Abbreviations

1	first person	NEG	negative
2	second person	NOM	nominative
3	third person	NP	noun phrase
ACC	accusative	OBJ	object
ADJ	adjective	OBL	oblique
AGR	agreement	P	preposition
CP	complementizer phrase	PASS	passive
Cp_{inf}	infinitival complementizer phrase	PCP	Pronominal Complex Predicate
DAT	dative	PL	plural
DEF	definite	PP	prepositional phrase
DO	direct object	Pred	Predicate
DOM	differential object marker	PRF	perfect
DP	determiner phrase	PRO	pronoun
EPC	external possessor construction	PROG	progressive
EPP	extended projection principle	PRS	present
EZF	ezafe	PST	past
F	female	REF	
FPR	full-form pronoun	S	subject
GEN	genitive	SBJV	subjunctive
IMP	imperative	SG	singular
IND	indicative	SOV	subject-object-verb
INS	isntrumental	Spec	Specifier
IO	indirect object	SupHighApplP	Super High Applicative Phrase
IP	inflectional phrase	TOP	topic
LF	logical form	TP	tense phrase
LV	light verb	TS	topic-shift
M	male	V	verb
N	noun		

CHAPTER 1 - INTRODUCTION

“It is the pervading law of all things organic and inorganic, of all things physical and metaphysical, of all things human and all things superhuman, of all true manifestations of the head, of the heart, of the soul, that the life is recognizable in its expression, that form ever follows function. This is the law.”

Architect Louis Sullivan, in "The tall office building artistically considered," Lippincott's Magazine (March 1896)

Idiosyncrasies in linguistic forms surfacing in the morphosyntax layer of grammar are often correlated with a distinct pragmatic/semantic function. Analyzing idiosyncratic forms requires an extensive study, ranging from diachronic considerations to studies in the syntax-semantics interface. It also requires analyzing a large set of data in order to be able to suggest an account that would capture all the data that demonstrate a particular idiosyncratic form.

This work is an attempt to suggest a revised analysis of a peculiar syntactic pattern observed in a growing class of complex predicates in colloquial Persian. Taking ideas from the whole range of previous literature, this analysis is based primarily on the theories proposed by Givón (1976), Anderson (2005), Levin (1993), Kulikov (2011), and Haig (2008). The purpose of this study is to propose a fine-tuned analysis of this irregular syntactic form in Persian, so that it could capture all such forms and the final result could be unified with and get subsumed by the wider description of complex predicates in this language, contributing to the diversity of periphrastic constructions in Persian described by previous literature (Bonami and Samvelian 2009).

1.1 The issue

Pronominal complex predicates, henceforth referred to as PCP for the ease of reference, are a sub-type of complex predicates in Modern Persian which seem to behave idiosyncratically with respect to agreement. It is also the only predicate type in Modern Persian where using a pronominal enclitic is obligatory, which is the reason I have coined this label for them.

Although Persian is a scrambling language, it underlyingly uses an SOV syntactic pattern. It shows regular subject-verb agreement, and by virtue of that, it is a pro-drop language. Table 1-1 shows the personal endings (i.e. subject agreement markers) used in Colloquial Persian.

Table 1-1: Personal Endings*

	Singular	Plural
1	-am	-im
2	-i	-in
3	Ø / -e	-an

So the simple verb *xordan* [1] 'to eat' is conjugated in the past tense as illustrated in Table 1-2.

Table 1-2: Simple past conjugation of *xordan* 'to eat'

	Singular	Plural
1	xord-am	xord-im
2	xord-i	xord-in
3	xord-Ø	xord-an

* In 3SG, null personal ending is used in the past tense and '-e' is used in the present tense.

[1] Infinitive form is composed of past stem of the verb, suffixed by "-an". So in *xordan*, *xord* is the past stem.

Modern Persian has a limited number of simplex verbs (250-300). These verbs have been and still are gradually eliminated by creating/substituting periphrastic expressions and compound verbs (Windfuhr, 1979). The conjugation paradigm of a compound verb is illustrated in Table 1-3 for the compound verb *dust dâštan* 'to like', in the present tense[2].

Table 1-3: Simple present conjugation of the compound verb *dust dâštan* 'to like'

	Singular	Plural
1	dust dâr-am	dust dâr-im
2	dust dâr-i	dust dâr-in
3	dust dâr-e	dust dâr-an

In PCP construction, however, the personal endings are absent. Instead, we observe affixation of object clitics to the non-verbal element of the complex predicate. Table 1-4 illustrates the paradigm of such a complex predicate for the 6 possible person/number combinations for the complex predicate *xoš umadan* 'to like' [pleasantness come] ((1-1) is the gloss for the first cell).

Table 1-4: PCP conjugation of *xoš umadan* 'to like'

	Singular	Plural
1	xoš=am umad-Ø	xoš=emun umad-Ø
2	xoš=et umad-Ø	xoš=etun umad-Ø
3	xoš=eš umad-Ø	xoš=ešun umad-Ø

[2] Conjugating in the present tense needs the present root of the verb. For the compound verb *dust dâštan*, *dâšt* 'had' is the past root, and the present root of it is *dâr* 'have'.

xoš=am umad-Ø

pleasantness=1SG come.PST-3SG.AGR (1-1)

I'm pleased. (Literally: pleasantness came to me.)

Table 1-5 is the paradigm of object clitics in colloquial Persian. Agnes Korn (2009) suggests that this paradigm represents a general oblique form of pronominal clitics which results from coalescence of genitive/dative and accusative clitics in Old Iranian.

Table 1-5: Object pronominal enclitics

	Singular	Plural
1	=am	=emun
2	=et	=etun
3	=eš	=ešun

These pronominal enclitics can be used interchangeably with full form pronouns in non-PCP contexts, and have various functions other than being used in PCP construction. They can function as the nominal argument of a noun, adjective, or preposition. They can also be used to express a nominal argument of the verb (Samvelian and Tseng 2010).

The glossed example (1-1) can optionally include a sentence-initial DP which shows the same morphosyntactic properties of the pronominal clitic suffixed to the preverbal element. Sentence (1-2) shows this optionality in a glossed format, and sentence (1-3) shows the same predicate with an added oblique argument realized in a PP, indicating the source of the feeling of 'pleasantness'. An interesting point that should be noted here is that co-occurrence of this clause-initial DP with a clitic with the same morphosyntactic properties (i.e. person and number) is not a violation of principle B of the Binding Theory.

(man) xoš=am umad-Ø

(I) pleasantness=1SG come.PST-3SG.AGR

I'm pleased. (Literally: pleasantness came to me.) (1-2)

(man) az Sara xoš=am umad-Ø

(I) from Sara pleasantness=1SG come.PST-3SG.AGR

I'm pleased by Sara. (Literally: pleasantness came to me from Sara.)
(intended: I feel good about Sara) (1-3)

(1-4) illustrates a sentence that is semantically similar to (1-3), but it uses the regular syntactic pattern in Persian, utilizing a compound verb as illustrated in Table 1-3.

(man) Sara.ro dust dâr-am

I Sara.DOM friend have.PRS-1SG.AGR

I like Sara. (1-4)

In contrast to sentence (1-4), which typifies the ordinary syntactic pattern of modern Persian (where verbs are conjugated by personal endings), constructions such as (1-2) and (1-3) are problematic in two respects. First, the verb is always in 3rd person singular form, regardless of the person and number values of the optional sentence-initial DP that seems to be the subject of the sentence. Second, the presence of an object pronominal clitic is obligatory in the clause, and eliminating it even in the presence of the sentence-initial DP would result in an ungrammatical sentence such as (1-5) which is a copy of (1-3) without the object enclitic pronoun.

* man az Sara xoš umad. (1-5)

1.2 Methodology

Although a diachronic overview of this construction is necessary for achieving a proper analysis, I will not go into the details of such an analysis since it's been accomplished by other linguists (cf. Arefi 2011). However, I will use some examples from Classical Persian to be compared against colloquial Modern Persian examples whenever I deem necessary in the data analysis chapter.

For the theoretical analysis, Levin's (1993) insights have been exploited at length, in conjunction with profound introspection to find semantic nuances in order to come up with a classification of predicates and find out potential resemblances among various situation types that use PCP frame. This process could lead us to learn about the function of this peculiar syntactic construction.

Since the language chosen in this study is colloquial Persian and a large spoken corpus was not available, the corpus I used for extracting these sentences was the World Wide Web. I used Google advanced search to find the data I needed, and I utilized the website "WebCorp: The Web as Corpus"[3] to count the tokens and types of the corpus I had collected.

First, I studied Levin's verb classes[4] and selected the verbs for which there was a Persian PCP construction. This investigation led to finding 120 PCP constructions (each semantically equivalent to a single verb). Then I tried to find 36 sentences for each PCP, including 18 sentences using the past root of the verb, and 18 sentences using the present root of the verb. Nevertheless, the number of detected sentences did not amount to 36 for

---

[3] http://www.webcorp.org.uk/

[4] Beth Levin's (1993) book, English verb classes and alternations, is an investigation of syntactic and semantic properties of English verbs. The basic assumption of this work is that the behavior of the verb with respect to its arguments is largely determined by its meaning. The systematic verb classification it provides helps with development of a lexical knowledge.

some PCPs. I realized that a few PCPs, e.g. *zannam bord*, *gamânam bord*, are mostly (more than 90%) used in the non-colloquial (standard/literary) register, but because of that 10% of colloquial occurrence, I kept them as part of the data to be analyzed. Ultimately, there were 110 PCPs left for which 2307 sentences were found and recorded in a Microsoft Excel file. This mini-corpus contains 26945 tokens and 6187 types, but since all the sentences are relative to this research, it could be counted as representative and it justly serves its purpose. Even if there is lack of data in the collected corpus to evaluate a hypothesis, the World Wide Web is always there as a reference to add to the corpus.

All the sentences recorded in the corpus were analyzed with regards to their argument structure. These analyses formed a basis for a statistical study which facilitated the classification of these verbs. Appendix-1 is a list of these 110 model PCP constructions. Based on this analysis and with recourse to Levin's insights in verb classification, I classified these predicates as represented in Chapter 3. This thorough examination helped me find parallels and patterns which led to a more fitting integration of theory and data.

1.3 Outline of the thesis

In Chapter 1, I have introduced the particular idiosyncratic syntactic frame that I have analyzed in this work. I have illustrated it with some examples, and have presented the idiosyncrasies it brings about.

In Chapter 2, I review the research previously done on this issue. The chapter is structured thematically, describing the views of various linguists on each topic. It constitutes of an introduction to the problem as studied in the previous literature, the diachronic studies, the description and explanation of verbal structure in this construction, the views on the issue of agreement and the function of pronominal clitics in these predicates.

Chapter 3 is a review of the theoretical backbone of the work and is structured thematically. The first section (section 3.1) is founded on Kulikov's "voice typology" (2011), explaining the concept of diathesis and diathesis alternation, and how it can affect the form of an expression. The second section (section 3.2) briefly defines complex predicates and periphrastic constructions in general, and examines the views on Persian complex predicates in particular (Samvelian 2012). The third section (section 3.3) reviews briefly the topic of subject agreement. It goes over Landau's (2003) views on the concept of subject and agreement, but it's mainly based on Givón's (1976) viewpoints on topichood and agreement. In defining topichood, Lambrecht's (1996) ideas are also exploited. The fourth section (section 3.4) looks at Anderson's (2005) opinions on pronominal clitics and their function and positioning in the clause. The last section (section 3.5) expounds on Haig's (2008) **external possessor construction** in Iranian languages. I argue in this section that the PCP construction is actually an instance of external possessor construction.

In Chapter 4, I provide a data-intensive analysis of PCP construction. In the first part of this chapter (section 4.2), I provide a lexical semantic analysis of the data, classifying verbs realized in PCP frame based on Levin's (1993) insights, guidelines and terminology. In the second part of the chapter (section 4.3), I provide an analysis based on the functions of light verbs and how they contribute to the semantic content of the whole complex predicate and the clause.

Chapter 5 is a wrap-up, applying the theoretical considerations provided in Chapter 3 to Persian PCP data classified in Chapter 4. I have formulated a novel proposal as an analysis for the PCP construction and have summarized how the reviewed theories apply to the problem at hand.

In Chapter 6, I have provided a summary of my analysis, and in section 6.2 I have proposed my recommendations for further research.

CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction

The construction introduced as **pronominal complex predicates** (PCP) in this study is the focus of heated debates in Iranian linguistics. A variety of analyses have so far been proposed to examine this construction in the interface of syntax and semantics. The construction has been labeled diversely by various linguists, which indicates the rainbow of viewpoints on this construction. Some linguists (Lazard 1992, Arjang-Sadeghi 1984, Fakhr Rohani 2001, Darzi-Tousi 2004, Najafian-Vahedi 2003, Vahedi 2006, Arefi 2011) have called them **impersonal constructions** (cf. Arefi, 2011). Najafian-Vahedi (2003) defined the impersonal constructions in Persian as any sentence that lacks a subject DP, where the verb always shows 3rd person singular agreement and there is no indication of a particular person. Then, the authors divide these constructions into four subcategories based on the way the verb is related to the other elements of the sentence. Among these subcategories, the third one is called the **templatic impersonal construction**, which is the same construction studied in this research. Arefi (2011) adopts the title **impersonal construction** based on this subcategory introduced by Najafian-Vahedi (2003). Karimi (2005) calls them subjectless constructions, distinguishing them from impersonal constructions which are known as **curtailed infinitives** by Persian grammarians and are called **short infinitives** by Karimi (2005).

Other linguists have described this construction as **nâgozar** (non-passing) (Khanlari, 1972), **Indirect (middle) verbs** (Windfuhr, 1979), **Compound Verbs of Experience**

(Barjasteh, 1983), **Psychological Predicates** (Sedighi, 2005), and **Enclitic Compound Verbs** (Rasekhmahand, 2010).

I do not criticize any of these labels since all of them have valid reasons behind them, but I postpone defending my new label to the concluding section.

In this chapter, a survey of the past literatures that have worked on this particular construction in Persian has been presented.

2.2 Diachronic Studies

Pronominal complex predicate (PCP), as defined and studied in this work, do not exist in Classical Persian. Instead, the experiencer surfaces as 'DP+râ' (Arefi 2011). In Classical Persian and today's literary Persian, 'râ' is the specific-oblique marker (Karimi, 1990), or more accurately the differential object marker. Based on the definition of oblique then, 'râ' can be used to mark any DP except those having nominative case[1]. The following are some examples ((2-1)[2], (2-2)[3]) of Classical Persian that are semantic equivalents of PCPs in Modern Persian.

xoš	âmad-Ø	soxan	šâh-goštâsb.râ	(2-1)
pleasantness	come.PST-3SG.AGR	utterance	King-Goštâsb.DOM	
King Goshtaasb liked the utterance.				

[1] This description of oblique is valid for a nominative-accusative alignment system.

[2] Shâhnâmeh

[3] Samak-e-ayyar

mâ.râ saxt gorosne ast-Ø (2-2)

we.DOM severely hungry be.PRS-3SG.AGR

We are very hungry.

Arefi proposes that the experiencer in these constructions is dative. She also mentions that dative experiencers are found in other languages as well, and suggests that this evidence supports the hypothesis that the experiencer in these constructions, whether in Classical Persian (DP+râ) or Modern Persian (clitic pronoun), is dative.

Khanlari (1986) examines this construction only in Classical Persian and does not compare it with its modern equivalent. He only describes this construction as not displaying subject-verb agreement at all, using an object clitic or full-form pronoun instead of personal endings. He describes this object clitic or full form pronoun as the logical subject of the sentence.

Abolghasemi (2006) (cf. Arefi 2011) suggests that in Classical Persian, whenever the logical subject is represented by an oblique clitic pronoun, or when there is no subject-verb agreement, it is actually a remnant of preterit transitive in Middle Western Iranian, which had had ergative case, i.e. the logical subject (agent) appeared with ergative case, and the verb agreement was with the logical object in absolutive case.

2.3 Verbal Structure

The fact that Persian enjoys a growing number of compound verbs has led many linguists to believe that the verb in PCP construction is also compound. However, there are also a number of prominent Iranian linguists who reject the compound status of the verb in PCP construction.

The first person who has studied PCP construction is Khanlari (1971). He acknowledges that the verb in PCP construction always expresses an “inactive” state or

condition and therefore calls them **stative verbs**. He believes that the verbs in PCP construction are compound, using six light verbs: *âmadan* 'to come', *budan* 'to be', *šodan* 'to become', *gereftan* 'to take', *bordan* 'to remove', and *zadan* 'to hit' (97-8).

Windfuhr (1979), who calls PCPs **indirect (middle) verbs**, proposes a tentative description in terms of case-grammar. He notes that in addition to the fact that the topic is always [+human] and corresponds to the obligatory dative experiencer realized by a pronominal suffix, what seems to be common in this construction is the presence of a SOURCE and a RESULT case. The source might not surface and remain implicit, but the result (i.e. the preverbal nominal element) is obligatorily expressed (127). He provides the following table of the sub-classes of these verb constructions based on Khanlari's data. In this table, (±) indicates the optional presence or absence of 'source', (–; +) indicates the optional presence with some members of the sub-class.

Table 2-1: Windfuhr's classification of PCPs based on Khanlari's data

(TOPIC)	(SOURCE)	RESULT	DATIVE	VERB
(+)	(–)	Indef. pronoun	+	'be/become'
(+)	(+; –)	Adjective	+	'be/become'
(+)	(±)	Adjective	+	'come'
(+)	(±)	Noun	+	'come/become'
(+)	(–; +)	Noun	+	'take/carry' (= 'overcome')

He also provides five examples of this construction, comparing them against their English and German translations[4] as he suggests based on his analysis (128). Each example is a representative of one row in Table 2-1, i.e. first row of the table is illustrated in example number 1, and so on and so forth.

[4] The translations provided for the five Windfuhr's example presented are exactly as suggested by him.

1. če-t-e? ‘what is the matter with you’ (German: ‘was ist dir?’)
   hiči-m nist ‘nothing is the matter with me’ (German: ‘mir ist nichts’)
2. sard-am-e ‘I am cold’ (German: ‘mir ist kalt’)
   (az ân) bas-am-e ‘I’ve had enough (of it)’ (German: ‘es reicht mir’)
3. (az ân) xoš-am âmad ‘I liked it’ (German: ‘es gefiel mir’)
4. (az ân) qahr-emân šod ‘we began to quarrel (because of that)’
   (az ân) heyf-am âmad ‘I was sorry (for that)’ (German: ‘es begann mir leid zu tun’)
5. xâb-am gereft/bord ‘sleep overcame me’
   (az ân) xande-am gereft ‘laughter overcame me (from that)’

Sentences (2-3) through (2-11) show the glossed version of Windfuhr’s examples.

če=t=e?
what=2SG=be.PRS.3SG.AGR
What’s the matter with you?

(2-3)

hiči=m nist-Ø
nothing=1SG NEG.be.PRS-3SG.AGR
Nothing is the matter with me.

(2-4)

sard=am=e.
cold=1SG= be.PRS.3SG.AGR
I feel cold.

(2-5)

(az ân) bass=am=e.
from that enough=1SG= be.PRS.3SG.AGR
I’ve had enough (of it).

(2-6)

(az ân) xoš=am âmad-Ø

from that pleasantness=1SG come.PST-3SG.AGR

I liked it.

(2-7)

(az ân) qahr=emân šod-Ø

from that quarrel=1PL become.PST-3SG.AGR

We began to quarrel (because of that).

(2-8)

(az ân) heyf=am âmad-Ø

from that pity=1SG come.PST-3SG.AGR

I was sorry (for that).

(2-9)

xâb=am gereft-Ø /bord-Ø

sleep=1SG take/carry.PST-3SG.AGR

Sleep overcame me.

(2-10)

(az ân) xande-am gereft-Ø

from that laughter=1SG take.PST-3SG.AGR

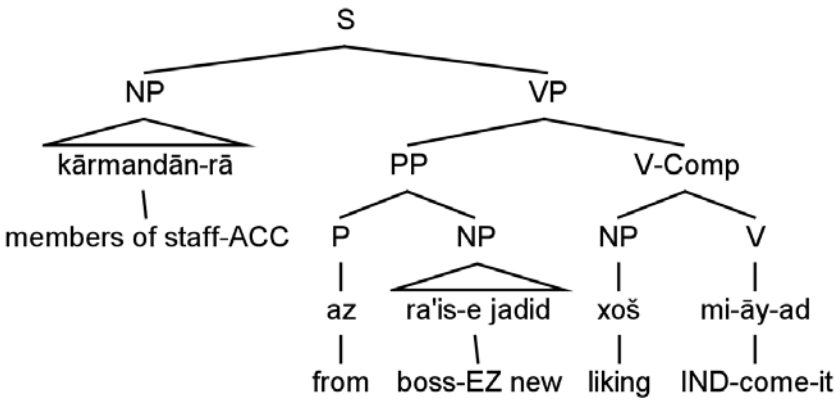
Laughter overcame me (from that).

(2-11)

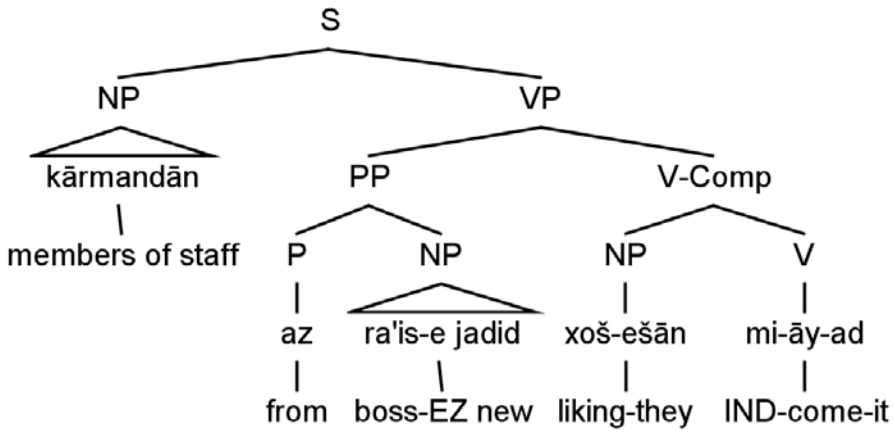
His tentative proposal is that these constructions imply an 'it' comparative to German 'es ist mir kalt', in which case the noun would appear in the predicative, and the construction would have the general case-frame: [(TOPIC) – 'it' – (SOURCE) – RESULT – DATIVE – Verb] (Windfuhr 1979: 128).

Barjasteh (1983) provides a full account of the history and current state of compound verbs in Persian (242-245), describing characteristics of compound verbs (245-326). He calls the PCP construction **compound verbs of experience**, i.e. a subcategory of compound verbs in which the nominal component denotes a human experience or a perceptual faculty (327). He analyzes them as derived lexically rather than syntactically. He also notes that many nominals with the semantic feature of [+human experience]

cannot be used in this construction (331). For example, *sardard* 'headache' is a [+human experience] nominal expression, but it cannot be used in a PCP frame. He assumes that the underlying structure of compound verbs of experience in Modern Persian is the classical equivalent of them where the experiencer surfaces as 'DP+râ'. He bases this assumption on two separate facts. First, Persian has an optional rule of "Accusative Object Cliticization", and this clearly shows that NP subjects in compound verbs of experience are underlyingly accusative. Second, the modern structure has evolved directly from the old correlates by application of a cliticization rule. Based on this assumption, he concludes that the syntactic rule of "feature copying" applies to these structures, paralleled with undergoing an obligatory syntactic rule of "Accusative Subject Cliticization". He illustrates this process with some examples one of which follows here. The trees (2-1) and (2-2), respectively, illustrate the underlying and surface structures (to use Barjasteh's words) of the sentence 'the staff like the new boss' using this construction as illustrated by Barjasteh (335-6). In fact, (2-1) is what (2-2) would look like in Classical Persian which Barjasteh assumes to be the underlying form. By applying the obligatory "accusative subject cliticization" rule proposed by Barjasteh, (2-1) changes to (2-2).



(2-1)



(2-2)

The above trees show the accusative subject cliticization obligatory syntactic rule. The accusative subject NP, *kārmandān-rā*, loses its accusative case, hence the differential object marker, *rā*, is dropped. The obligatory accusative subject cliticization rule has cliticized the nominal element of the compound verb. Barjasteh says that the (2-1) is the underlying structures of the sentence, but he doesn't mean the deep structure in the sense of transformational syntax. It is actually only one step before the surface structure, before the obligatory accusative subject cliticization rule is applied.

In his famous paper "Compound Verbs in Persian", Dabir-Moghaddam (1997) discusses the PCP construction very briefly. He rejects Barjasteh's (1983) account as "totally unfounded", claiming that they are not compound verbs at all, but "fully-fledged sentences in which the nominal element is the subject and the obligatory rule of subject-verb agreement in Persian systematically treats these nominals as the subject" (45). He notes that these constructions can use compound verbs in place of simple verbs in sentences such as (2-12)-(2-14).

sar=am dard gereft-Ø
head=1SG pain take.PST-3SG.AGR
'I got a headache.' (2-12)

hosele=am sar raft-Ø
Patience=1SG head go.PST-3SG.AGR (2-13)
'I became impatient.' (Lit. my patience overflowed.)

nafas=am band âmad-Ø
Breath=1SG closure Come.PST-3SG.AGR
'I was out of breath.' (Lit. my breath stopped.) (2-14)

Dabir-Moghaddam believes that PCPs are actually "frozen sentences whose verb meaning is metaphorically extended." (46)

Vahidian-Kamyar (2003) calls PCP construction **fixed-agreement verbs** since verbs are always in the 3rd person singular form in them, and the subject agrees with a pronominal clitic. He has identified 14 verbs that are used in this construction: *âmadan* 'to come' -implying "inclination", *oftâdan* 'to fall', *bâlâ âmadan* 'to rise', *bar-dâštan* 'to pick up' –"where the pronominal clitic is suffixed to 'bar', i.e. the preverbal preposition", *bordan* 'to remove', *budan* 'to be', *dar-âmadan* 'come out', *raftan* 'to go', *zadan* 'to hit', *šodan* 'to become', *kardan* 'to do/to make', *kešidan* 'to pull', *gereftan* 'to take', *nešastan* 'to sit'.

Like Barjasteh, he also explains that these verbs cannot co-occur with just any nominal and there is a restricted set of nominals that can be used with these verbs in this construction. For that reason, he recommends that they should be mentioned in dictionaries as collocations. He also provides a long list of nominals that commonly co-occur with these light verbs.

Sedighi (2005) calls PCPs **psychological constructions**. She agrees with Dabir-Moghaddam (1997) about the non-compound status of these predicates. She enumerates ten features for these constructions and analyzes the predicate as a vP projection. She

proposes that these are actually full propositions to which a Super High Applicative head[5] might optionally be added. Like Dabir-Moghaddam, Sedighi believes that the psychological state, known by others as the non-verbal part of the compound verb, is in fact the theme argument of the unaccusative light verb which moves to the subject position. Therefore, the regular subject-verb agreement is also maintained since the psychological state is a non-human nominal element, hence being inherently 3rd person singular.

2.4 Agreement/Clitic

If, like many analyses, the optional sentence-initial DP in PCPs is treated as subject, then the construction would lack the regular subject-verb agreement, which would be ungrammatical in Persian. As stated in the introduction chapter, the verb in these constructions is always realized in 3rd person singular form, while the clause-initial DP which seems to be the subject of the sentence could have any person and number values. Below are the various analyses that have been proposed about this issue.

Dabir-Moghaddam (1997) believes that regular subject-verb agreement is maintained in PCPs, and the fact that the verb is always realized in 3rd person singular form is because the subject that it agrees with is actually the state occurring to the experiencer.

---

[5] Sedighi (2005) introduces the notion of Applicative head, following Pylkkänen (2002), as "a syntactic head whose function is to license an argument DP (applied argument) which is not a true argument of the verb, divided based on their semantic and syntactic properties into High and Low Applicatives. The applicative head in high applicatives denotes a thematic relation between an individual and an event, and in low applicatives it denotes a transfer of possession relation (cf. Sedighi 2005, 123-130). Sedighi adopts Pylkkänen (2002) and proposes a new category of applicatives, calling it the **Super High Applicative** head, which takes an event complement (like high applicatives), but it's different from high applicatives in that it has a Tense requirement. This requirement causes the Super High Applicative head to constitute a strong phrase.

He thinks that the full DP experiencer that could optionally appear sentence-initially is actually not subject, but does not elucidate his idea any further.

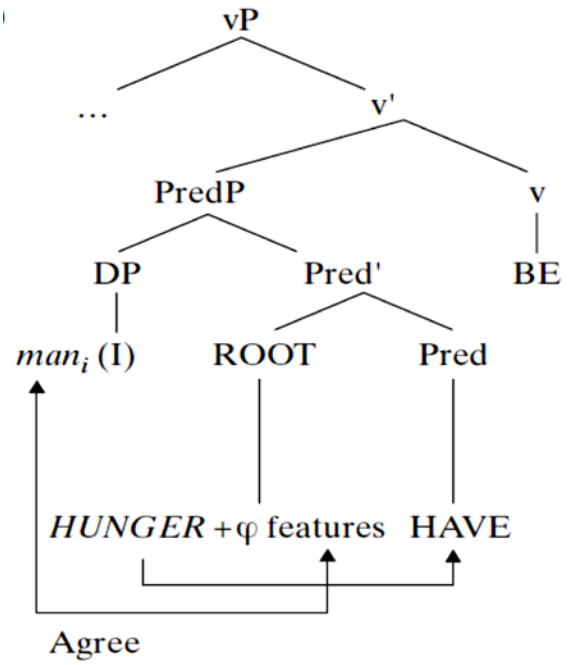
Rasekhmahand (2010) calls this construction **Enclitic Compound Verbs**. Supposing that the experiencer is subject, he suggests that these constructions lack the regular pattern of subject-verb agreement in Persian. However, he believes that in these constructions the pronominal enclitics have been grammaticalized and are now subject agreement markers. He provides an example from another phenomenon in colloquial Modern Persian where we observe a potential reanalysis of pronominal enclitics to agreement markers. This new function which is occasionally used in colloquial Persian and it is not yet obligatory is the growing use of 3SG pronominal enclitic functioning as a 3SG personal ending in preterit verbs which are conjugated with null-morpheme personal ending. He notes that this new addition has been developed to compensate for the non-canonicity of the paradigm of preterit verbs in Persian.[6]

Karimi (2005) argues that the clause-initial optional DP cannot be the grammatical subject, because it does not morphologically agree with the verb. She calls PCPs **subjectless constructions** and recognizes them as complex predicates. She categorizes these constructions into two subgroups: *Inalienable Possessor Constructions* and *Inalienable Pseudo-Possessor Constructions*. In the *Inalienable Possessor Construction*, she interprets the semantics to be possessive, and compares it to 'J′ai friod' experiencer-type constructions in Romance languages. She suggests that these sentences have an underlying possessor construction containing *HAVE*, where the obligatory clitic that is co-indexed with the optional overt DP in the clause-initial position is the *possessor*. She argues that the copy of the possessor DP which is represented by the person/number

[6] 3rd person singular verb in the past tense is conjugated with a null morpheme, making it syncretic with the past stem of the verb. This new optional stylistic addition in the colloquial register substitutes the null morpheme with the 3SG object clitic in order to compensate for the non-canonicity of the paradigm.

features (ϕ-features), surfaces as a clitic pronoun on the noun. She illustrates the structure of the sentence (2-15) in the following tree (2-3).

(man$_i$) gorosne-am$_i$-e (2-15)
I hungry-me-is
I am hungry.



(2-3)

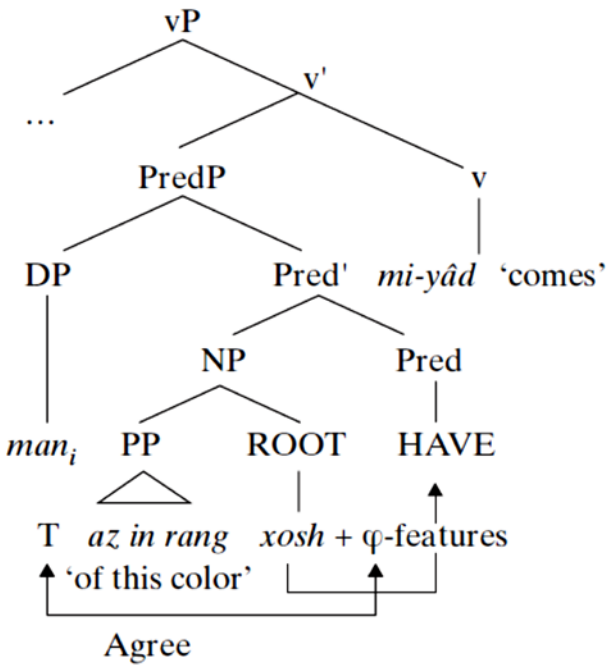
Karimi suggests that when the overt DP is missing, *pro* is in the possessor position (81), and the clitic pronoun which is actually the copy of the possessor is reminiscent of a general double clitic construction that exists in Persian.

In an *inalienable possessor construction*, the light verb is always "BE", while in the *inalienable pseudo-possessor constructions*, the choice of light verb is different, and the complement of the light verb has a more complex internal structure, but we still have a *'HAVE'* interpretation. The tree (2-4) represents the structure of the sentence (2-16) as analyzed by Karimi (84).

man$_i$ az in rang xoš-am$_i$ mi-â-d

I from this color pleasure=1SG IND-come.PRS-3SG.AGR (2-16)

I like this color.



(2-4)

She proposes that the (nominal) root is incorporated into *HAVE* and the light verb is spelled out. The construction lacks a grammatical subject, and the optional DP is actually the topic of the sentence.

In agreement with Dabir-Moghaddam (1997), Sedighi (2005) states that the psychological/mental state is an element which satisfies the EPP feature of v by moving to the specifier position of the verb, and it induces agreement on it. She notes that the psychological/mental state is inherently 3rd person singular, and this explains why the verb is always spelled out in this form. About the obligatory presence of the pronominal clitic, she explains that the main construction is complete without the presence of a full

DP experiencer in the initial position of the sentence in which the experiencer is obligatorily encoded as a clitic pronoun on the psychological/mental state.

Regarding the clitic, she believes that it is not the argument of the verb, but the argument of the theme/psych state. The theme subject contains an applied argument which is interpreted as the possessor/recipient/location/affectee of the theme (172). She suggests that this applied argument (clitic) is not directly involved in the event, but it is related to it through a possession relation with the psych state (173).

Lazard (1992) notes that the so-called 'clitic pronouns' are actually personal suffixes. He asserts that this personal suffix represents the interested person, i.e. the person who finds himself/herself in the state of, or being subject to, the event described by the phrase (111).

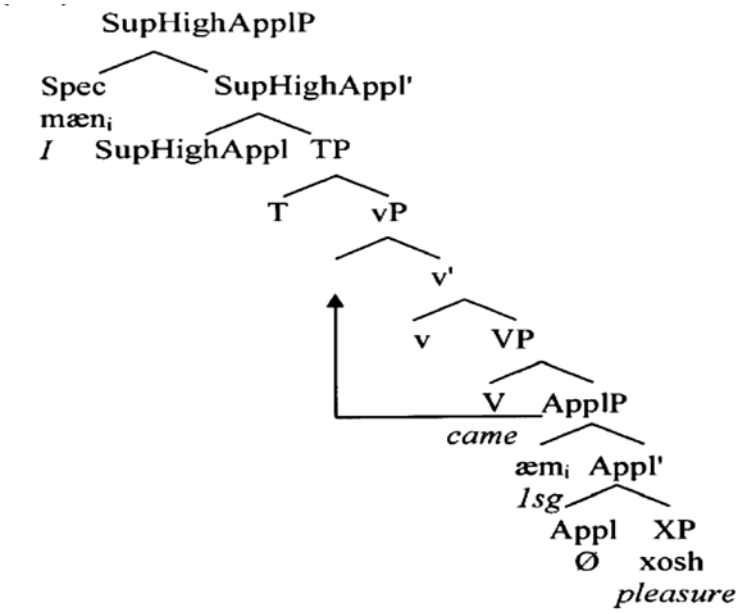
2.5 Subject/Topic

Sedighi (2001) particularly scrutinizes the optional sentence-initial DP in this construction. First she attempts to compare them with "quirky subjects". Quirky subjects are subject-like DPs which have a non-nominative case and a non-agent semantic role, but show some properties of the subject. Sedighi proves that the optional DPs in these constructions are neither quirky subjects, nor even subjects. As a matter of fact, they are "left-dislocated constituents", and the subject in these sentences is the psych state, not the subject-like experiencer DP.

In her PhD dissertation, Sedighi (2005) revises her past analysis, arguing that the optional sentence-initial experiencer DP is in fact an applied argument. She proposes that the Tense requirement in psych constructions imposes the structure of a *Super High Applicative* projection (175) above TP which licenses the experiencer in its specifier (174). When an overt experiencer is not present sentence-initially, this projection is still present and contains a [+human/+mental state] feature which is phonologically null. She

argues that there is a difference between the forms with or without the overt experiencer DP, and that difference is in the information communicated (175). She shows that in case of the presence of an overt experiencer DP, the sentence has a topic with discourse functions (such as quantification and wh-formation); and in case it is absent, the sentence is topic-less. She believes that the overt experiencer DP, i.e. the topic, is a non-core applied argument, licensed by an applicative head, and merged/base-generated in a higher position.

Sedighi's analysis is inspired by Shibatani's account of dative subjects. He believes that the DP is assigned to a complete clause and is co-indexed with an element within the clause. Sedighi's analysis of the syntactic structure is illustrated in (2-5).



(2-5)

Applying an argument (here, the experiencer) to a complete clause or a sentential predicate requires the argument to be recognized/co-indexed within the clause, making the argument and the clitic, which is realized on the theme nominal, internally connected through the Super High Applicative phrase. The clitic behaves like a bound variable as opposed to a full form pronoun, so unless there is a binder the sentence would be

incomplete. Hence, even in the absence of the sentence initial DP, a referential *pro* would be present which would be co-indexed with the pronominal enclitic.

CHAPTER 3 – THEORETICAL CONSIDERATIONS

Persian PCP construction is a periphrastic construction (or complex predicate) which shows a particular type of construal of the event by the speaker. This construction shows idiosyncrasy in agreement pattern compared to regular subject-verb agreement observed in Persian. Another idiosyncrasy is that PCP construction demands obligatory presence of an object pronominal enclitic on the non-verbal element of the complex predicate.

In order to analyze such a construction which shows such idiosyncrasies, we need to be familiar with the theories that could explain such a construction. For that purpose, we should review some theories about complex predicates, diathesis alternation (which imposes valence alternation), and the notion of agreement in general since it could help in coming up with an analysis for the observed idiosyncrasy. On the other hand, the obligatory presence of the pronominal enclitic needs explanation, which would necessitate a review of the theories on pronominal clitics.

The other important notion that I should introduce in this chapter is the **external possessor construction** proposed by Haig (2008), which is a construction observed in Iranian languages and PCP construction is an instance of it. So it would definitely be crucial to have at least a brief introduction to it.

3.1 Valence Alternation

According to the framework developed by Leningrad/St. Petersburg Typology Group, there are two main levels of representation of linguistic structure (Kulikov 2011). The first level is the level of semantic roles which is determined by the semantic class of verbs. The second level is the level of grammatical relations that is responsible for realizing arguments in the clause. There are three major formal devices for encoding grammatical relations, including case marking, verbal agreement, and word order. These three parameters together determine the syntactic structure of the clause. In some cases, only one of these parameters is at work to determine syntactic functions.

The notion that is closely related to the concept of valence and is determined in terms of the two mentioned levels of representation is the notion of *diathesis* which is a pattern of mapping semantic arguments onto grammatical relations. The basic or neutral diathesis is the most common, unmarked way of representing an event. For example, the neutral diathesis for a simple transitive verb is:

Semantic Argument Level (role)	X (Actor)	Y (Undergoer)
Syntactic Function Level (case)	S (NOM)	DO (ACC)

Any modification of diathesis is reflected in changes of syntactic patterns. A popular well-studied diathesis alternation is passivization which results in:

X	Y		X	Y
S	DO	➔	Obl / -	S

In languages with relatively strong case-marking (e.g. Sanskrit), diathesis alternation could be reflected in changes in case-marking. There are also languages that encode passive diathesis by a special verbal morpheme. In simple cases, passivization suggests no changes in the semantics of the sentence. In this sense, passive, antipassive, dative shift and some other diathesis alternations are categorized as diathesis alternations *senso*

stricto, and alternations such as causative, anticausative, benefactive and some other diathesis alternations are categorized as diathesis alternations *senso latiore* (cf. Kulikov 2011:4). Syntax, semantics and morphology all play an important role in defining and describing different voices[1].

There are many cases where a group of similar diatheses is represented by the same morphological form of the verb. This is called a *diathesis cluster*, and the members of the cluster normally share some features. One of these clusters is "Middle Voice" which might include passive, conversive, anticausative, reflexive, reciprocal, antipassive and subject version (cf. Kulikov 2011:25). Kemmer (1993) quotes Lyons (1969: 373), describing middle voice as occurring when "an action or state affects the subject of the verb or his interests". Central to the notion of middle voice is the affectedness of the subject. Crosslinguistically, middle voice has the semantic property of having an affected chief nominal participant (Kemmer 1993: 8). The name "middle" appropriately describes the intermediate state of these verbs in transitivity, being somewhere between one-participant and two-participant events.

The middle diathesis imposes a particular form of expression since all the propositions that are middle are of a similar situational type (Talmy, 1972). In other words, they include real-world information filtered through the conceptual apparatus of the speaker. Middle voice, like all other diatheses, significantly relies on how the language user construes a real-world situation in different ways (Givón, 2001).

Persian PCP construction uses two sub-types of the middle diathesis alternation including conversive and anticausative. So I will explain these two alternations since they are relevant to the construction studied in this work.

[1] Voice is a regular encoding of diathesis through verbal morphology.

3.1.1 Conversive

Conversive is a type of diathesis alternation that occurs when the semantic distance between the main two arguments is less than an 'Actor-Undergoer' pair. In other words, the affectedness of the undergoer is less than in the agent-patient case. In this case, the initial subject may degrade less severely than in passives, becoming an indirect or oblique object of relatively high rank. Conversive often happens to verbs of perception and emotional states, i.e. mental events, which subcategorize for a stimulus and an experiencer argument. An example from Russian follows (Kulikov 2011:12). The dog is the experiencer in both sentences. In (3-1) it is realized as a direct object showing accusative case, and in (3-2) it is realized as an indirect object showing genitive case.

X	Y	➔	X	Y
S	DO		IO/Obl	S

Grom ispuga-l-Ø sobaku
thunder:NOM frighten-PST-M.SG dog:ACC
'The thunder frightened the dog.'

(3-1)

Sobaka ispuga-l-a-s′ groma
dog:NOM frighten-PST-F.SG-REF thunder:GEN
'The dog was frightened by the thunder.'

(3-2)

3.1.2 Anticausative

Anticausative (decausative)[2] is a diathesis alternation that removes the agent from the structure.

X	Y	➔	~~X~~	Y
S	DO		−	S

Anticausatives and agentless passives use similar morphological markings crosslinguistically. This is because they both promote the patient argument and demote the agent argument. Some languages, like Russian, use distinct morphological markings for these two categories.

a)
Ivan razbi-l-Ø vazu
John:NOM broke-PST-SG.M vase:ACC
John broke the vase.

b)
Vaza razbi-l-a-s’ (*Ivanom)
vase:NOM broke-PST-SG.F-REF (John:INS)
The vase broke (*by John).

c)
Vaza by-l-a razbi-t-a (Ivanom)
vase:NOM be-PAST-SG.F broke-PART.PRF.PASS-SG.F (John:INS)
The vase was broken (by John).

In cases where there is no morphological distinction between the two, they can be distinguished only by semantic criteria: in passives, the existence of an agent is implied

[2] It is called ‘inchoative’ by Haspelmath (1993).

even if it is not overtly expressed, while in anticausatives, the situation could come about spontaneously (Comrie 1985: 326). Distinguishing these two is one of the most complicated problems a linguist might confront in a syntactic study of verbs (Kulikov 2011: 24).

3.2 Complex (-or Periphrastic-) Predicates

Ackerman and Webelhuth (1998) define *predicate* as the association between a lexeme, a morphosyntactic property set, and a form. If the form of a predicate is a word, the predicate is called *synthetic*, and if the form is a sequence of words linked by syntactic relations, the predicate is called *periphrastic* or *analytic* (Bonami and Webelhuth 2013).

Complex predicates are defined as predicates composed of more than one grammatical element, each contributing a non-trivial part to the information hidden in the complex predicate (Alsina, Bresnan and Sells 1997). In complex predicates, the argument structures of syntactically independent elements are brought together by some kind of argument fusion mechanism different from the usual types of complementation. It is rather some kind of co-complementation (Alsina and Butt 2008). The nominal element of the complex predicate is selected the same way other arguments of the verb are selected, but it is different from other arguments of the verb semantically. It constructs a predicate with the verb, like a lexical verb.

Modern Persian has about 250-300 simplex (i.e. monolexemic) verbs. Putting aside the unfamiliar and unused verbs, it amounts to about 100 verbs the most frequent of which could function either as a lexical (heavy) verb or a light verb. As a result, Persian makes heavy use of constructions that fall under the general category of complex predicates, in order to compensate for the small size of the repertoire of simplex verbs. Bonami and Samvelian (2009) assert that Persian complex predicates have three functions. They can be used to express idioms, copular constructions, and productive and

compositional light verb constructions. PCP construction is of the last type, i.e. it is a light verb construction which is compositional as much as non-PCP complex predicates in Persian are, and productive in particular usages which would be discussed in section 4.3.2.

There are arguments on whether complex predicates are formed in the lexicon or syntax. Goldberg (2012) believes that Persian complex predicates are formed in the lexicon since they can undergo infinitival nominalization. Samvelian (2012), on the other hand, believes that they are syntactic sequences. Samvelian's argument seems to be more acceptable and well-founded. There are a number of criteria that would determine whether a complex predicate is formed in the lexicon or syntax. If formed in the lexicon, a complex predicate would have a single phonological stress, which is true about Persian complex predicates, both in PCP and non-PCP forms. It can be put to morphological derivation, which is generally not the case with PCP construction[3] but true about non-PCP complex predicates. However, as Samvelian prudently mentions, this process of morphological derivation is not limited to complex predicates and it actually can be used with just any object-verb syntactic sequence that has undergone incorporation and has lost the differential object marking. So, infinitival nominalization or any other type of morphological derivation is not a sufficient condition for considering that the complex predicate is not formed in the syntax. The other criterion that makes us judge the complex predicate as being formed in the lexicon is that it is inseparable except with affixes and clitics. This is definitely not the case with Persian complex predicates, whether in PCP or non-PCP form.

[3] There are some examples of nominalized PCPs found in the corpus, but they are considered to be marked. The examples found were among the most frequently used PCPs, e.g. *xoš umadan* 'to like', hence it is possible that the infinitival nominalization of regular complex predicates has been over-generalized by a few native speakers, not paying attention to the markedness of such gerunds.

Samvelian (2012) points out that if a complex predicate in Persian is formed in the syntax, it would be separable by affixes, by the auxiliary verb *xâstan* 'to want', and by prepositional phrases. The second point to consider is to look at the preverbal element. If it's a noun, it could be modified by an adjective, quantifier, or determiner. The following three glossed sentences (3-1, 3-2, 3-3) are examples of PCPs (retrieved from the corpus) where the verbal and non-verbal element of the complex predicates are separated by a prepositional phrase (stimulus).[4] These show that Persian complex predicates are syntactic sequences, i.e. they are formed in the syntax rather than in the lexicon.

hers=am	az=at	gereft-Ø.
anger=1SG	from=2SG	take.PST-3SG.AGR

I'm pissed off by you.

(3-3)

ahl-e	siâsat	nist-am	o	xoš=am	az=aš	ne-mi-â-d.
fan-EZF	politics	NEG.be.PRS-1SG.AGR	and	pleasure=1SG	from=3SG	NEG-IND-come.PRS-3SG.AGR

I'm not a fan of politics and I don't like it.

(3-4)

da'vâ=m	bâhâ=š	šod-Ø	badjur.
quarrel=1SG	with=3SG	become.PST.3SG.AGR	badly

I quarreled with him/her badly.

(3-5)

In non-PCP complex predicates in Persian, there is a prevalent situation of a posteriori compositionality (Samvelian 2012). The predicates with high frequency are more likely

[4] The English translations do not go into the nuances of meaning here since the purpose of bringing these examples is only to show they are separable.

to be a core semantic model based on which the semantically similar themes could compose a complex predicate with a similar form, i.e. using the same light verb. As a result, there would be parallels between form (syntax) and meaning (semantics). Therefore, a semantic classification of similar predicates would help to extract the rules behind their compositionality.

The small size of the repertoire of simplex verbs in a language and the polysemous situation it causes makes it almost impossible to draw a bold line between lexical and light verbs in a language like Persian with about a hundred frequently-used monolexemic verbs. By the same token, most complex predicates are semi-lexical, which is somewhere in the middle of the spectrum of being lexical and syntactic. There seems to be a lexicalization cline for complex predicates, starting from purely syntactic and compositional and changing toward getting more and more lexicalized. This change seems to be correlated with the frequency of use of a certain complex predicate; the more a complex predicate is used, the more lexicalized it gets. The complex predicates on the lexicalized end would be used as a model for producing novel complex predicates that are semantically similar to the more lexicalized complex predicate.

3.3 Subject Agreement

What is a subject? Is there a fixed set of properties that can be attributed to subjects? Evidently, subjecthood is a “range”, and any clause in any language has at least one noun phrase with certain conditions and relations that makes them eligible to be a subject (Boeckx 2000). There is no single syntactic position in a clause to which a noun phrase qualified for being a subject could be mapped (Harley 1995).

> The past 20 years have seen the gradual deconstruction of the notion “subject” (McCloskey 1997, Sigurðsson 2000). Subjecthood is no longer viewed as a package deal; rather, particular subject properties are distributed over separate dimensions (structural positions, case, agreement, EPP, thematic roles, topicality, etc). Importantly, again and again we see that these properties can be dissociated,

> within and across languages, such that the question "Is X a 'real' subject?" becomes increasingly vague.
>
> (Landau 2003:80-1)

Experiencers tend to show anomalous behavior crosslinguistically. Since they can occur both as subjects (with theme objects) and objects (with theme subjects), they usually pose difficulties for the thematic hierarchy and argument linking (Adger 2006, Croft 1993). Landau (2003) proposes that when the experiencer is surfaced as an object, it is certainly a subject by LF. The theme argument is *overtly* raised to the specifier of TP, and the experiencer is raised to a second [Spec, TP] *at LF* (83).

If as Landau (2003) proposes, the theme argument is overtly raised to the specifier of TP when the experiencer is surfaced as an object (while it is a subject by LF), this could lead to a situation with two layers of agreement, or as Kim (2004) calls it, *hybrid agreement*, which allows for a morphosyntactic agreement and a semantic agreement within the same sentence (Sharifian 2007).

In his paper “Topic, Pronoun and Grammatical Agreement”, Givón (1976) deconstructs the notion of grammatical agreement by rejecting the “myth” that verb agrees with the *subject*, and reconstructs this notion by proposing that the verb agrees with *topic*. As a matter of fact, a language might diachronically reanalyze the topic argument as the normal subject or object of the neutral (non-topicalized) sentence pattern. When this happens, subject-topic agreement is automatically reanalyzed as subject-agreement, and object-topic agreement is reanalyzed as object-agreement. This is hand in hand with the re-analysis of pronouns as agreement morphemes. Givón also “lays to rest” the implicit postulation that agreement and pronominalization are two different processes, and suggests that they are essentially the same phenomenon.

Before elaborating more on Givón’s theory of agreement, I deem it necessary to devote a subsection to clarifying what is meant by a topic.

3.3.1 Topichood

I take Lambrecht (1996) as the reference to get to know the idea of topichood, what it is, and what it is not. The concept of *topic* as he examines refers to *sentence topic* or *clause topic* which is related to the grammatical form of the sentence, as opposed to *discourse topic* which is more related to text cohesion. He also distinguishes topic from the element which comes first in the sentence, since this element could either be a topic or a focus. He adopts the definition of topic by referring to the definition of subject "in traditional grammar", yet having in mind that these two notions cannot be conflated. "The topic of a sentence is the thing which the proposition expressed by the sentence is *about*." This definition of topic which is established on the pragmatic notion of *aboutness* has the consequence of being relative since *aboutness* is an inherently vague notion. The question "what this sentence is about" can never have one and only one answer. This means that topichood is a range and there are degrees to which different elements in a proposition could qualify as topic (119). Such a pragmatic-based definition of topic will sometimes make it impossible to determine the topic of a sentence solely on the basis of syntactic structure. By the same token, topics and grammatical subjects are not necessarily the same.

In the theory of information structure, the term "old information" (which is interchangeable with the term "pragmatic presupposition" or simply "presupposition") is defined as "the set of propositions lexicogrammatically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered". On the other hand, the term "new information" (which is interchangeable with the term "pragmatic assertion" or simply "assertion") is defined as "the proposition expressed by a sentence which the hearer is expected to know or take for granted as a result of hearing the sentence uttered". It should be noted that what is intended by "knowing" a proposition, is "having a mental representation of the denotatum" of that proposition.

The definition proposed for *topic* makes it inherently related to the notion of *presupposition*. "Topic is the matter of current concern about which new information is *added* in an utterance." A proposition about a topic is construable only if the referent of the topic is "under discussion" or otherwise available from the context (150).

The best way to find discourse-pragmatic categories such as topic is to find and analyze *allosentences*, i.e. alternative sentence structures expressing the same proposition. Although the morphosyntactic and prosodic structure of individual sentences can be analyzed without recourse to the categories of information structure, it is only information structure that can explain the difference between *allosentences* (120).

3.3.2 Topic Agreement

Givón (1976) proposes that verb agreement is governed by the universal hierarchy of "topicality". The basis of this hierarchy is the likelihood of an NP argument to be the topic of a sentence. He lists this hierarchy as:

a) Human > non-human
b) Definite > indefinite
c) More involved participant > less involved participant
d) 1^{st} person > 2^{nd} person > 3^{rd} person

It's only natural that human conversation, particularly in colloquial speech, is anthropocentric, and even egocentric. This anthropocentric nature of discourse results in item (a) of the above topicality hierarchy, and the egocentrism of it makes (d) indisputable. Item (b) reflects the definition of *topic* as being old information, hence definite. Item (c) leads to a hierarchy of grammatical cases based on the likelihood of being topic.

Agent > Dative > Accusative

Based on this hierarchy, if a sentence lacks an agent argument, it will search for a dative argument to identify it as the sentence topic. If the search fails and the sentence does not have a dative argument as well, the accusative argument will take the lead and be identified as the sentence topic.

The common idea that the verb agrees with subject is only the consequence of the highly universal pairing of "topic" as a discourse function with "agent" as a semantic role into the grammatical relation "subject". Furthermore, the fact that agents and datives are often [+human] is another justification for this topicality case hierarchy. [+human] arguments are normally [+definite], so line (b) of the topicality hierarchy is compatible with topicality case hierarchy since agents and datives are more frequently [+human] compared to accusatives.

Givón suggests that in "subject prominent" languages, subject NP holds most of the topic functions, which makes evident why subjects are the most frequent arguments to develop grammatical agreement. The subject is highly qualified for being a topic, so it undergoes a topicalization (topic-shift) process, which co-occurs with anaphoric pronominalization. Over time, speakers decide that the construction is too marked, so they re-analyze it as neutral (topicless) syntax, with the topic-subject getting re-analyzed to merely a subject, and the topic-agreement anaphoric pronoun getting re-analyzed as subject-agreement. Givón calls this process "de-marking". The pronoun inevitably undergoes a process of cliticization and gets morphologically bound to the verb.

TS ('Marked')				Neutral (Re-analyzed)	
The man,	he	came	→	The man	he-came
TOP	PRO			S	AGR

Reinhart (1995) suggests that natural languages have the following tendencies:

(a) Subjects tend to be topics

(b) Objects tend to be focus
(c) Definites tend to be topics
(d) Indefinites tend to be focus

She proposes that whenever a language goes against one of these tendencies, some sort of marking will be required. Thus, an indefinite topic subject or a focus subject will need special marking as will a definite (no-focus) object or topic object (cf. Ghomeshi 1997).

Given the topicality hierarchy (Givón 1976), in a situation where the highest likelihood of being a topic does not belong to the subject of the sentence but to the object, we will have the object topicalized. According to Reinhart (1995), topic objects will need special marking. The same "de-marking" process Givón proposes for the re-analysis of subject-topic to subject and topic-agreement anaphoric pronoun to subject-agreement marker is at work with a topic object, the anaphoric pronoun (re-analyzed as an agreement marker) being the special marking suggested by Reinhart. This will ultimately lead to object agreement after re-analysis has occurred.

Universally, languages with a viable subject-verb agreement system would more readily allow subjectless sentences in anaphoric contexts, hence being known as pro-drop languages. This could be true about sentences with object agreement as well, since being pro-drop is related not to dropping the subject, but to dropping the topicalized NP which is marked in the sentence by an agreement marker.

Givón (ibid) concludes that grammatical agreement is fundamentally a topic related phenomenon that arises from anaphoric pronominalization in topical discourse contexts. He asserts that this is the only viable explanation for the diachronic rise of grammatical agreement from pronouns.

3.4 Pronominal Clitics vs. Morphological Affixes

What is interesting for syntacticians in studying clitics is basically examining the category of clitic pronouns. For these researchers (such as Kayne (1975) and his followers), it's not the phonological features of the clitics that is exciting. What fascinates them is a significant place where non-clitic elements do not appear. In this respect, clitics have a morphosyntactic sense (Anderson 2005).

The classic method for distinguishing clitics and affixes is using Zwicky and Pullum's (1983) set of tests. Here is a summary of them.

a. Clitics have a low degree of selection with respect to their hosts; affixes a high degree of selection.
b. Affixed words are more likely to have accidental or paradigmatic gaps than host + clitic combinations.
c. Affixed words are more likely to have idiosyncratic shapes than host + clitic combinations.
d. Affixed words are more likely to have idiosyncratic semantics than host + clitic combinations.
e. Syntactic rules can affect affixed words, but not groups of host + clitic(s).
f. Clitics, but not affixes, can be attached to material already containing clitics.

Zwicky (1977) divides the clitics into simple and special groups. Simple clitics or phonological clitics have deficient phonological form in that it lacks prosodic structure at the level of the prosodic word (Anderson 2005: 23). Special clitics or morphosyntactic clitics are "linguistic elements whose position with respect to the other elements of the phrase or clause follows a distinct set of principles, separate from those of the independently motivated syntax of free elements in the language" (Anderson 2005: 31). Klavans (1982, 1985) provides a descriptive typology of special clitics, noting that any given special clitic is located with respect to some domain with which it is syntactically and semantically associated. Based on this domain-based description, she suggests three

types of clitics. *Sentence clitics* which take scope over the whole sentence, *NP/DP clitics* associated with nominal expressions, and *phrasal clitics* which can be associated with any phrase type. So special clitics can be categorized based on the syntactic domain within which they are located (CP/IP, DP/NP, XP) (Anderson 2005: 79).

Klavans also proposes two more parameters about the positioning of clitics. First, the orientation of the clitic with respect to the first or last element in the domain in which it is located. This is called *dominance* by Klavans and *anchoring* by Anderson. The second parameter specifies the clitic's placement as *before* or *after* the anchoring element (hence *proclitics* and *enclitics*).

With respect to such a characterization of clitics, there are three types of clitics amply attested in natural languages: initial clitics, second-position clitics, and final-position clitics. Sometimes the clitic anchoring is not to the first or last element within the domain, but to the *head* of that domain.

In order to account for special clitics in an explicit grammar, the popular picture is a syntactic view where the special clitics are taken to act like lexical items, being introduced into the sentence and then moved to their surface position. Anderson rejects this view as unsuitable and takes special clitics to be morphological objects, i.e. overt morphological markers of the morphosyntactic properties of a phrase (Anderson 2005: 83). According to this view, clitics are *phrasal affixes*. Like morphological affixes, clitics have a rather fixed order, and they are essentially never free (while syntactic elements are). In this regard, they behave more like morphology than like syntax (Anderson 2005: 84).

Pronominal clitics are, in the minds of many linguists, the canonical examples of special clitics. Anderson (2005:227) assumes that pronominal clitics are the functional morphology of phrases. Generative grammar looks at pronominal clitics as certain types of pronouns occupying an argument position just as just as any other nominal expression, and that they move to their position in surface structure. Such an analysis, Anderson

suggests, is rejected by evidence from languages with *clitic doubling*, since in these languages they are *not* mutually exclusive with overt nominals. Consequently, Anderson proposes a different view where instead of movement of pronominal clitics from some argument position to the surface form, they are an "overt reflection of the properties of that position, construed as part of the functional content of the clause and realized by a principle of phrasal morphology as a modification of the phonological form of the clause" (Anderson 2005: 229). He further suggests that pronominal clitics are similar to verbal agreement markers.

Anderson explains two types of agreement:

1) Some set of features from an argument are copied to the head, hence the morphosyntactic representation of the head will contain these features. These features can trigger "the introduction of overt morphological markers in the derivation of appropriate word form" (Anderson 2005: 232). This type of agreement is called "registration" and has no further syntactic consequences.

2) Genuine agreement which involves both co-indexing and feature identity. It is a process where the morphosyntactic representation of the verb should be co-indexed with an argument position to which the verb assigns θ-roles. The index is referential, and co-indexation takes place with the agreeing argument position. Such a definition of genuine agreement has syntactic consequences. Unlike in registration, appearance of an overt nominal expression with which the verb (and hence the clause) is co-indexed via Agreement is a violation of the Binding Conditions, because the nominal would be bound within its clause (Anderson 2005: 234). Therefore, R-expressions and full form pronouns cannot appear in such positions if we have Agreement in the clause (and not merely registration).

However, Anderson explains that phonologically null pronominal elements (*pro*) are not like overt nominal expressions and presence of them in the agreeing position does not pose a problem to the binding principle since *pro* is not referential, i.e. it does not have an

index. So he suggests that *pro* is not a real pronoun, but just an empty category which has its own restrictions like other empty categories (ibid).

Anderson proposes a typology of the relations between agreement and arguments it indexes based on these considerations.

1) Languages like English and French are one end of the spectrum, having a non-coindexing registration-type agreement, thus never requiring or allowing argument positions to be empty in the absence of other specific displacements or deletions.

2) *Pro-drop* languages such as Italian or Persian can have an optionally empty agreed-with position. The optionality is within the operation of Agreement which always copies the features of the argument, but the co-indexation is optional. In case of co-indexation, there would be a pro; and where there is no co-indexing, an R-expression or full form pronoun appears.

3) In some languages (like Pemon [Cariban, Venezuela]), realization of an argument could be *either* by an overt nominal *or* by verbal agreement, *but not both*. In such languages, agreement is referential but optional. The agreement marker is in complementary distribution with a full form nominal argument. In this sense, co-indexing *agreement* can be called *strong* agreement, and non-coindexing *registration* can be called *weak* agreement.

3.5 External Possession Construction

In his book "Alignment Change in Iranian Languages", Haig (2008) extensively studies what he calls **external possession construction (EPC)** in Iranian languages. He reviews the nature of the Old Persian Genitive and describes its function as including a quite broad but not fully arbitrary range of semantic roles. He recognizes Benefactive to be at the core of the functions of Genitive with the other attested functions as radial extensions from this meaning. He notes that Genitives cannot readily be classified based

on semantic roles. So it is difficult or nearly impossible to have a unique interpretation of the semantic role of an argument case-marked as Genitive.

He indicates that although the function of the Genitive could be described as a disjoint list of semantic functions, such a description is missing some significant generalizations for a number of reasons. First, such a list would be remarkably similar to the list of functions covered by Datives in a number of languages (particularly Indo-European languages), so such a list would not be purely arbitrary. Second, particular examples of the Genitive cannot be distinctly attributed to a single semantic role, so it might be the case that categorizing the Genitives based on semantic roles is not really relevant at all. So Haig suggests that rather than treating the Genitive as a list of disjoint and arbitrary semantic roles, it would be wiser to come up with a more coherent analysis.

He introduces the concept of **indirect participation** (as opposed to direct participation) as a coherent analysis that captures the function of the Genitive. At the core of the notion of indirect participation lays the Benefactive semantic role which is a sentient being, who is neither Agent nor Patient, but whose interests are affected by, or perceived to be affected by, the event expressed in the verb (Haig 2008:58). The notion of indirect participation is not readily compatible with analyses that work on a one-dimensional scale of control and affectedness, such as the scale of the thematic relations suggested by Van Valin (2001:31):

Agent > Experiencer > Recipient > Stimulus > Theme > Patient

But the notions of affectedness and control are not necessarily inversely related with some semantic roles such as benefactive or experiencer. In order to compensate for this incompatibility, Lehmann et al. (2004:7-9) proposes two dimensions of participation: direct and indirect. What necessitates the introduction of the second dimension, i.e. indirect participation, is the fact that events cannot be defined solely in terms of the participants directly involved in an event, and the participants for whom the event is, in one way or another, relevant should also be taken into account. The characteristic

representative of the dimension of indirect participation is the Benefactive, and it is not surprising that many languages have almost grammaticalized devices for expressing this role. By the definition of indirect participation, it is only natural that it is prototypically restricted to sentient beings, most commonly humans, who have subjective perception, i.e. a cognizance of personal interests. The notion of indirect participation in Iranian languages, Haig suggests, would be captured with the feature combination [+volitionality, -instigation, +affectedness]. Næss (2007) notes that this feature combination is typically displayed by the so-called "Datives" cross-linguistically. This domain generally covers the roles of recipients, benefactives and experiencers, with a fairly frequent extension to possessors (Næss 2007: 199). Figure 3-1 is a simplified overview of the semantic roles situated with respect to the two dimensions of direct and indirect participation.

Lehmann et al. (2004) note that the participant roles in the center of the model "become indistinct", i.e. one could consider a single role as either direct or indirect. In this model, a bundle of semantic roles are defined by the dimension of indirect participation, and the Old Persian Genitive covers the area characterized by "maximum indirect participation", which clearly does not include roles such as instrument or patient, both of which are direct participants.

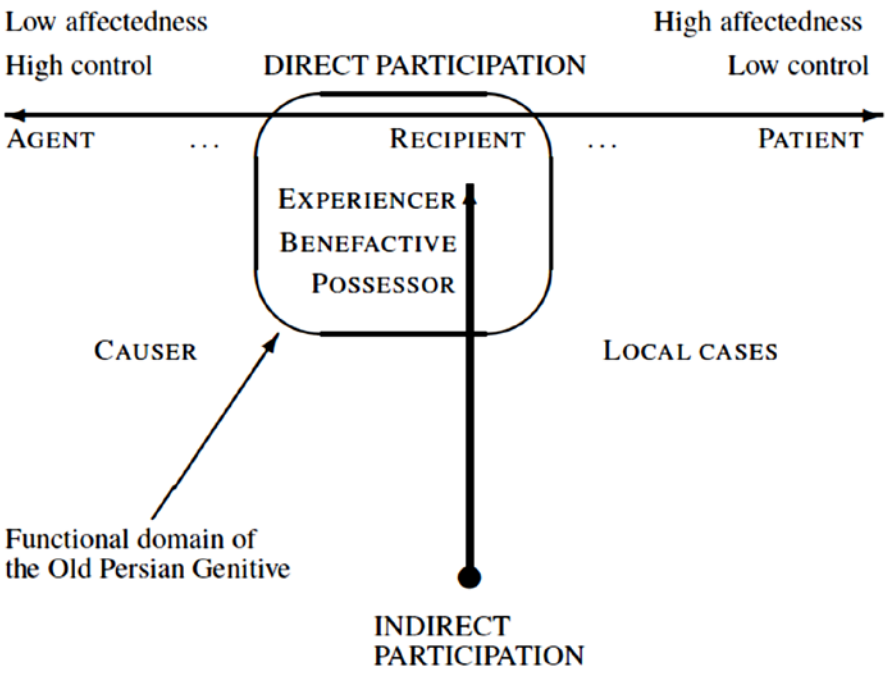


Figure 3-1: The function of the Old Persian Genitive

Cross-linguistically, it is common that indirect participation is expressed using a single case. This single case is commonly Dative which often functions as coding of an external possessor. As Haspelmath (1999) explains, in an **external possessor construction (EPC)**, a possessive modifier occurs NP-externally as a constituent of the clause. This is syntactically crucial since the possessor is a clause-level constituent rather than a sub-constituent of a NP. The semantic map of the functions typical of Dative case, as provided by Haspelmath (1999:126), covers categories including Predicative Possessor, Direction, Recipient/Addressee, Experiencer, Benefactive, External Possessor and Judicantis. This list, Haig suggests, could be applied "verbatim" to the Old Persian Genitive. So the Old Persian Genitive was functionally exhibiting a profile very similar to the Datives found in numerous modern Indo-European languages.

Haig views EPC as a type of construal made available by the semantics of indirect participation. So the term EPC is used to describe those instances of clause-level Genitives which can be interpreted as possessors. Syntactically, however, there seems to

be no distinction between clausal genitives which express external possessor, or benefactives, or experiencers. He assumes that in Old Iranian, the Genitive/Dative also coded the possessor in an EPC, where the possessor reading is only "one construal of the broader category of indirect participation".

The construction studied in this work, i.e. PCP construction, is an instance of Haig's EPC which is observed in Old Persian (and even more generally, in Old Iranian). The functional domain of PCP overlaps with the functional domain of Old Persian Genitive which includes a bundle of semantic roles, all of which could be described as having indirect participation. As Haig suggests, EPC is a type of construal of an event or state. Such a view is compatible with the notion of diathesis. The domain of indirect participation is like a cluster, with a more or less similar diathesis or construal. For expressing such a cluster, as noted in section 3.1, a language would employ a similar morphosyntactic device. Old Persian used EPC to encode indirect participation, and PCP construction is an instance of it which is used today in Modern Persian.

CHAPTER 4 – DATA ANALYSIS

4.1 Introduction

The construction I call Pronominal Complex Predicate (PCP) has been treated variously in the past literature. In some of them (Karimi 2005, Dabirmoghaddam 1997, Barjasteh 1983, Sedighi 2005, Windfuhr 1979, Rasekhmahand 2010) the range of the data reviewed does not cover all the predicate types using this syntactic pattern. Arefi (2011) accounts for a wider range of data, but still does not provide a thorough detailed analysis that covers her full range of data. Her analysis is mainly diachronic, and this is probably the source of its inadequacy.

I endeavored in this work to investigate a quite large body of data consisting of sentences using PCPs, and I further tried to use Levin's (1993) classification of English verbs as a model for classifying the (somewhat) periphrastic verbs in the PCP construction. In a class-based construction of verb lexicon, the central assumption is that the syntactic frames are a direct reflection of the underlying semantics. By classifying verbs we can come up with generalizations about verb behavior (Kipper, Dan and Palmer, 2000). Verb classes are hierarchically organized in a way that all the members of a class are semantically similar and have a common set of semantic roles, syntactic frames, and diathesis alternations (ibid).

By putting to practice the notion of "diathesis alternation", Levin's lexical semantic work corroborated the theoretical notes I have provided in the "diathesis alternation" section in chapter 3. I used Lambrecht's (1996) notion of "allosentences" to ensure the

verity of my hypotheses, and to find the patterns of mapping of semantic arguments onto syntactic functions.

The first part of this chapter (section 4.2) is devoted to data intensive lexical semantic analysis of the PCP construction. In the second part (section 4.3), I apply the relevant theoretical considerations introduced in Chapter 3 to these constructions, proposing an analysis that explains almost all the idiosyncrasies attributed to them.

4.2 Lexical Semantic Analysis of the Data

4.2.1 Possessor Centered Self

In these PCPs the pronominal clitic is a genitive experiencer, but not the subject or even the direct object of the sentence (it is merely a possessor). All of them are the result of the underlying form undergoing a preposition drop alternation[1].

4.2.1.1 Verbs of Memory

Class Members: *yâd raftan* 'to forget' [memory go], *yâd umadan* 'to remember' [memory come], *yâd budan* 'to remember; to have in my mind' [memory be]

It is easy to detect the shared element in these verbs: *yâd* 'memory'. Verbs of memory are complex predicates underlyingly composed of 'PP + V', where the DP (object of preposition) has the form '*yâd*=ϕ', with the morphosyntactic properties (i.e. person and number) of the experiencer realized in a pronominal clitic form. In the prepositional

[1] This alternation, as proposed by Levin (1993), is found with certain verbs of motion that take directional phrase complements. These verbs are found intransitively with a directional phrase or transitively with a path or goal (or in a few instances a source) as direct object. The transitive frame appears to be derived from the intransitive frame by "dropping" the preposition (Levin 1993, 43-44).

form, the clitic could be easily replaced with a full form pronoun or a DP. This substitution can happen in PCP frame too, but it's very rare. It should also be noted that according to our data, the prepositional form is mainly used in the standard/literary register and is often avoided in colloquial forms. The two sentences in (4-1) are allosentences, the first one in the underlying prepositional form and the second one in the preposition-dropped PCP form.

az	yâd-e	man	raft-Ø.	yâd=am	raft-Ø.
from	memory-EZF	1SG	go.PST-3SG.AGR	memory=1SG.GEN	go.PST-3SG.AGR
I forgot it. (Lit. It went out from/left my mind.)				I forgot it.	

(4-1)

As a perceptual capacity possessed by humans, a person's memory is viewed as a locus and a concrete or abstract object moves to and from it or resides in it, creating various verbs of memory by using different verbs of motion. According to the corpus data, this object is a nominal argument which could be a DP (e.g. I forgot the lesson), an infinitival CP (e.g. I forgot to do the homework), or a finite CP (I forgot that we should have done the homework by today). Another case is when verbs of memory occur with a null (*pro*) argument which is co-indexed with a contextually salient argument. In other words, a topic (which is old information and commonly present in the consciousness of both sides of the dialogue) is not mentioned for the purpose of economy because it is perceived as shared knowledge by the speaker.

The following table shows the variety of semantic roles *yâd* (the experiencer's memory) can take, what verbs would accompany each predicate, what is the literal denotation of the predicate, and what prepositions each take in the underlying form, i.e. in the alternate form with a preposition.

Table 4-1: verbs of memory

locus	θ-role of 'yâd'	Preposition	Verb	Gloss	Literal meaning
yâd	Source	az: from	raft	went/left	It went from (left) my mind.
yâd	Goal	be: to	umad	came	It came to my mind.
yâd	Goal	be: to	oftâd	fell	It fell (in)to my mind.
yâd	Location	dar: in	mund	stayed	It stayed in my mind.
yâd	Location	dar: in	bud	was	It was in my mind.

Sentence (4-2) is a glossed example from the collected corpus.

ne-mi-dun-am	či	šod-Ø	in	yâd=am	umad-Ø
NEG-IND-know.PRS-1SG.AGR	what	become.PST-3SG.AGR	this	memory=1SG.GEN	come.PST-3SG.AGR

I don't know what happened that I remembered this.

(4-2)

In sentences (4-3) and (4-4), a case is shown which is neither prepositional (since there is no preposition in the sentence), nor PCP (since the verb is not in the 3rd person singular form). These are actually cases with a clause-initial *pro* which is simply not spelled out. This can be realized by trying to topicalize an argument in the sentence. In these cases, the topicalized element would not be the experiencer (the possessor of memory), but the object that moves to or from the experiencer's memory, which is both the logical and grammatical subject of the sentence. Hence there is no idiosyncrasy in such sentences and it uses neutral syntax. The thing that makes them similar to PCP construction is that they are verbs of memory using a pronominal clitic and no preposition, but the verb is not in 3SG. The point is that such sentences are really prepositional forms, but the preposition has been dropped merely for economy purposes. This means they are intransitive sentences with a locative adjunct (taking the memory as a location), while the verbs of memory using PCP construction are transitive, taking DP or CP (finite or infinitival) arguments.

har vaqt yâd=am oftâd-i, bexand-Ø

any time memory=1SG.GEN fall.PST-2SG.AGR smile.IMP-2SG.AGR

Whenever you remembered me, smile!

(4-3)

emšab yâd=am umad-i

tonight memory=1SG.GEN come.PST-2SG.AGR

Tonight I remembered you. (Lit. You came to my mind.)

(4-4)

The motion verb is a heavy verb in prepositional forms and a light verb with a rather small degree of bleaching in PCP forms. The prepositional forms are in neutral diathesis, showing the neutral syntax of the language. The preposition drop alternation (proposed by Levin) results in a PCP construction. The experiencer is interpreted as having more volition or direction of attention to the stimulus in PCP constructions (Croft 2011). This alternation is caused by a shift in diathesis, the result being a change in the mapping of semantic onto syntax.

4.2.1.2 Verbs of Desire

Class Members: *meyl kešidan* ‘to desire (to do something)/crave (something)’ [desire drag]; *ešq kešidan* ‘to will (to do something); to want (to do something) at will’ [love drag]

It is easily detectable that the shared element in these complex predicates is the verb *kešidan* ‘to pull/to drag’. The experiencer’s desire and will are dragged toward a desired object, action, or state. The desired theme could be a DP (when the experiencer craves

X)[2] or an infinitival CP (when the experiencer wills/desires "to do" something). Among the 18 example sentences in our corpus that contain the PCP *meyl kešidan*, four have contextually-salient objects of desire (i.e. stimulus), seven have a DP stimulus (i.e. the craving sense of the verb is used), and seven have a CP_{inf} stimulus. For the PCP *ešq kešidan* however, there are four sentences having a contextually-salient stimulus, and 14 having a CP_{inf} stimulus. On the other hand, the crave sense of *meyl kešidan* has an alternant form with a preposition. The glossed examples (4-5), (4-6) and (4-7) illustrate these alternating forms.

torši ziâd meyl=am mi-kesh-e
pickles very crave=1SG.GEN IND-drag.PRS-3SG.AGR
I crave pickles too much!

(4-5)

meyl=am ziâd be torši mi-kesh-e
crave=1SG.GEN very to pickles IND-drag.PRS-3SG.AGR
I crave pickles too much! (Lit. my desire is dragged toward pickles too much!)

(4-6)

meyl-e man ziâd be torši mi-kesh-e
crave-EZF 1SG.FPR very to pickles IND-drag.PRS-3SG.AGR
I crave pickles too much! (Lit. my desire is dragged toward pickles too much!)

(4-7)

Again, based on Levin's preposition drop alternation analysis, the verb in the frame with a PP complement is intransitive and the PCP frame is transitive. Note that the sentences (4-5) and (4-6) show that in the prepositional intransitive frame, the experiencer could be realized in either clitic or full form, but the first sentence does not have a counterpart without a non-clitic experiencer. Also, in the first sentence, *torši*

[2] As clarified in the class members, it's only 'meyl kešidan' that could mean crave and therefore can take a DP.

'pickles' could optionally be marked with *ro* (the differential object marker) and is the object argument of the verb. Here again, the change is due to a diathesis modification similar to the one in verbs of memory.

4.2.1.3 Conjecture Verbs

Class Members: *šakk bordan* 'to get suspicious' [suspicion carry], *zann bordan* 'to get suspicious' [suspicion carry], *gamun bordan* 'to guess' [guess carry], *hads bordan* 'to conjecture' [conjecture carry]

Conjecture Verbs (Levin 1993) are a type of perception verb (Croft 1993) where the experiencer is a bit more qualified as a subject because it is engaging in an activity over which it has some volition or control. It's the experiencer that initiates the action in "suspecting" or "doubting". Nevertheless the experiencer is still only a possessor in the perspective adopted in the PCP frame.

Four conjecture PCPs were found in the corpus, all using the light verb *bordan* 'to carry', with the preverbal elements being *šakk* 'doubt', 'suspicion', *zann* 'suspicion', *gamun* 'guess', *hads* 'conjecture'. The second and third ones were found only in the literary register. The fourth one had too few results to be considered in the analysis. So the only representative of conjecture verbs remained to be analyzed was the PCP *šakk bordan* 'to get suspicious' [suspicion carry].

Like the verbs 'think', 'wonder', or 'consider', in addition to the experiencer subject argument, these verbs take a CP_{inf} or PP (to+DP) complement as well. The DP and CP_{inf} are semantically *stimulus* or *cause*. Using the light verb *bordan* 'to carry' which is a motion verb indicates a motion from experiencer's mind (where suspicion happens) toward the suspicious object/cause of suspicion. As Croft (1993) suggests, in these verbs the stimulus is often a governed oblique because it is not affected by the action of the experiencer. This is supported by our data where 64.28% of the sentences have a governed oblique stimulus (88.88% CP_{inf} and 11.11% DP), 21.43% have a cause known

to the hearer through sentence-external context, and 14.28% have a stimulus expressed through an adverbial clause of cause.

4.2.2 Dative Centered Self

Dative is defined as “the ultimate object toward which the predicate is directed” and “the most salient participant after the subject”. The semantic difference between dative and accusative complements rests on “a more stative vs. a more dynamic meaning” of the verb, where the accusative complement of the verb is affected by the action more directly than the dative (Van Belle and Van Langendonck, 1996).

In ditransitive structures, the main usage of dative could be a) a material transfer (e.g. to give), b) a verbal and perceptual transfer (e.g. to tell), c) a physical motion (e.g. to take), or d) an abstract motion (e.g. to subject). In intransitive structures, the main usage of dative could be a) interest (to obey), b) physical movement (to arrive), c) psych movement (to please), and possessive dative: ‘he has a headache’ (ibid, xi).

The intransitive possessive dative has certain selectional restrictions, with the set of verbs using them being limited to those expressing a physical or emotional sensation. Some examples from Spanish verbs admitting dative possessives are the verbs *coçar* ‘to itch’, *doer* ‘to ache’, and *tremer* ‘to tremble’ (ibid, 143).

The majority of PCP constructions in colloquial Persian are intransitive constructions with a dative centered self. In this section, we review the various types of PCP constructions containing a dative centered self.

4.2.2.1 Possessive Dative

Dative Possessive (*Dativus Possessivus*) is a type of possession famously found in Latin and a number of Romance languages, typically using a copula instead of the verb ‘to have’. In colloquial Modern Persian, questioning and answering about one’s age is a

predicate type syntactically similar to a PCP, and it would be explained as being a possessive dative construction. The question and answer in (4-8) are examples of it.

čand	sâl=et=e?	panj	sâl=am=e
how many	year=2SG=be.PRS.3SG.AGR	five	year=1SG=be.PRS.3SG.AGR
How old are you?		I'm five years old.	

(4-8)

Like a PCP construction, the verb (copula) is always in 3rd person singular form, and the pronominal enclitic is suffixed to the preverbal nominal phrase and shows the same morphosyntactic properties of the person whose age is being expressed or asked about. On the other hand, there is another predicate type in Persian for expressing one's age that uses the verb 'to have' instead of the copula 'to be' (4-9).

čand	sâl	dâr-i?	panj	sâl	dâr-am
how many	year	have.PRS-2SG.AGR	five	year	have.PRS-1SG.AGR
How old are you?			I'm five years old.		

(4-9)

The other evidence that helps to analyze this predicate type is a little background knowledge of Classical/literary Persian. Sentence (4-10) is an example from 'Shams-e parande'[3], a contemporary play written in literary style, and sentence (4-11) is an example from Classical literary Persian taken from 'Divân-e Shams' by Rumi, each followed by its "have-verb" equivalent.

[3] Shams-e parande (the flying Shams), a play about Rumi's life by Pari Saberi

ma-râ	yârâ-(y)e	goftan	nist-Ø,	to-râ	guš-e	šenoftan	hast-Ø?
1SG.FRP-DOM	vigor-EZF	saying	NEG.be.PRS-3SG.AGR	2SG.FPR-DOM	ear-EZF	listening	be.PRS-3SG.AGR

I don’t have the ability to say, do you have the ear to hearken?

(4-10)

yârâ-(y)e	goftan=am	nist-Ø,	guš-e	šenoftan=at	hast-Ø?
vigor-EZF	saying=1SG	NEG.be.PRS-3SG.AGR	ear-EZF	listening=2SG	be.PRS-3SG.AGR

I don’t have the ability to say, do you have the ear to hearken?

yârâ-(y)e	goftan	na-dâr-am	guš-e	šenoftan	dâr-i?
vigor-EZF	saying	NEG-have.PRS-1SG.AGR	ear-EZF	listening=2SG	have.PRS-2SG.AGR

I don’t have the ability to say, do you have the ear to hearken?

goft-Ø	ân-če	yâft	mi-na-šav-ad	ân=am	ârezu-st-Ø
say.PST-3SG.AGR	that-which	find	IND-NEG-become.PRS-3SG.AGR	that=1SG	wish-be.PRS-3SG.AGR

S/he said: the one that is not found, I desire that (I have a desire for that.)

(4-11)

goft-Ø	ân-če	yâft	mi-na-šav-ad	ân=râ	ârezu	dâr-am
say.PST-3SG.AGR	that-which	find	IND-NEG-become.PRS-3SG.AGR	that=DOM	wish	have.PRS-1SG.AGR

S/he said: the one that is not found, I desire that (I have a desire for that.)

So the possessive could be expressed in two forms: either using the verb ‘to have’ with the verb agreeing with the possessor (contemporary), or using copula (to be) in the 3SG

form, and the possessor realized either as an object enclitic on the possessed entity or as an oblique full-form pronoun suffixed by the differential object marker 'râ'[4].

Another common predicate that could be analyzed under the possessive dative sub-category is the verb that Levin categorizes as a judgment verb: *edde'â šodan/budan* 'to claim' [claim become/be] (e.g. sentences (4-13) and (4-14)). The preverbal nominal element is *edde'â* 'claim', and there is an equivalent non-PCP format for this verb (sentence (4-12)), using the light verb *dâštan* 'to have' instead of *šodan/budan* 'become/be'.

man	edde'â	dâr-am	ke	mi-fahm-am
1SG.FPR	claim	have.PRS-1SG.AGR	that	IND-understand.PRS-1SG.AGR

I claim that I understand. (Lit. I have claim that I understand.)

(4-12)

man	edde'â=m=e	ke	mi-fahm-am
1SG.FPR	claim=1SG.GEN=be.PRS.3SG.AGR	that	IND-understand.PRS-1SG.AGR

I claim that I understand. (Lit. my claim is that I understand.)

(4-13)

man	edde'â=m	mi-š-e	ke	mi-fahm-am
1SG.FPR	claim=1SG.GEN	IND-become.PRS-3SG.AGR	that	IND-understand.PRS-1SG.AGR

I claim that I understand. (Lit. my claim becomes that I understand.)

(4-14)

[4] There is a controversy on what is 'râ' a marker of. The famous function is 'accusative marker', but clearly it is more generally an 'oblique marker'. Karimi (1990) characterizes it as a marker of obliqueness, specificity, and discourse functions. The least uncontroversial view is that it is a differential object marker.

4.2.2.2 Psych Predicates

Verbs of psychological state have been labeled differently by various linguists: psych-verbs (Levin 1993, Belletti and Rizzi 1988, Pollard and Sag 1992), emotive predicates (Brekke 1988), psychological predicates (Aijmer 1972), psych movement predicates (Postal 1971) to name but a few. They typically take two arguments: experiencer and stimulus (Levin 1993). Croft (1993) proposes that if a mental state is expressible as either a subject-experiencer or an object-experiencer form in a given language, the version with a subject-experiencer indicates more volition or direction of attention to the stimulus. In Yoruba (cf. Croft 1993: 66) this option exists, and the difference in meaning would be that "where the experiencer is the grammatical object, the emotion is thought of as coming on him of '**its**' own volition, as it were, while where the person is the subject, he is thought of as summoning up the emotion, which is entirely under his control".

Colloquial Persian has this option for some verbs. These verbs are actually complex predicates and the two versions usually use different light verbs. In the subject-experiencer version, the verb agrees with the subject. For example, the subject-experiencer version of the complex predicate *gerye kardan* 'to cry' for 1st person singular experiencer is *gerye kard-am* 'I cried', but the object-experiencer version of it is *gery=am gereft* 'I felt like crying'.

Based on Levin's verb classification, I will classify psych predicates which have a dative object experiencer and use the PCP syntactic frame, listing the most common PCPs with their general (i.e. not fully detailed) meaning. Later in this chapter, I will go into finer semantic details and draw some generalizations that would explain why the PCP frame is not restricted to a closed set of verbs, but is a quite productive template.

4.2.2.2.1 Admire Verbs

Class Members: *xoš umadan* 'to like' [pleasantness come], *bad umadan* 'to dislike' [bad come], *zur umadan/gereftan* 'to begrudge' [force come/take], *âr umadan* 'to feel abashed' [abashment come], *estefrâq gereftan* 'to feel like vomiting (because of being

exposed to a detestable stimulus)' [vomit take], *hers gereftan/umadan* 'to get peeved' [peeve take/come], *laj umadan/gereftan* 'to feel pissed-off' [irritation come/take], *oq umadan/gereftan* 'to feel like puking (because of being exposed to a detestable stimulus)' [puke come/take]

4.2.2.2.2 Amuse Verbs

Class Members: *xašm gereftan* 'to get angry' [anger take], *qeiz gereftan* 'to get enraged' [rage take], *vahšat gereftan* 'to get afraid' [fear take], *delšure gereftan* 'to get anxious' [anxiety take], *ta'ajjob gereftan/umadan* 'to get amazed' [amazement take/come], *tars gereftan/umadan* 'to get afraid' [fear take/come], *šarm gereftan/umadan/šodan* 'to get ashamed' [shame take/come/become], *xejâlat gereftan/umadan/šodan* 'to get embarrassed' [embarrassment take/come/become], *qorur gereftan* 'to get proud' [pride take], *qosse gereftan/šodan* 'to get sad' [sorrow take/become], *larz gereftan* 'to shudder' [shudder take], *dard gereftan/umadan* 'to feel pain' [pain take/come]

4.2.2.2.3 Marvel Verbs

Class Members: *xošk zadan/bordan* 'to be petrified' [stiff hit/carry], *boht zadan/gereftan/bordan* 'to get bewildered' [bewilderment hit/take/carry], *mât zadan/bordan* 'to get astonished' [astonished hit/carry]

4.2.2.2.4 Pity Verbs

Class Members: *rahm umadan* 'to pity' [mercy come], *heif umadan* 'to regret' [alack come], *hasudi šodan* 'to envy' [envy become]

All psych predicates have non-PCP neutral-diathesis allosentences which use a different light verb (or are simplex verbs like *tarsidan* 'to fear' for *tars gereftan* 'to get frightened' [scare take]). For example, *hers gereftan* 'to get peeved' [peeve take] is taken from *hers xordan* 'to get peeved' [peeve eat], and *oq gereftan* 'to feel like puking' [puke take] is taken from *oq zadan* 'to puke' [puke hit].

4.2.2.3 Long Verbs

Class Members: *havas umadan* ‘to crave’ [crave come], *xâreš gereftan* ‘to itch for’ [itching take]

There was very little data found for long verbs on the Internet. Nonetheless I included them here as a part of my analysis since it was one of Levin’s classes which had PCP-type equivalents.

4.2.2.4 Verbs Involving the Body

These verbs include natural involuntary body reflexes when the person experiencing them is considered to be more or less ‘afflicted’ by the reflexes (4.2.2.4.1). These verbs also include voluntary facial and vocal (but nonverbal) expressions in reaction to an external or internal stimulus (4.2.2.4.2). Snooze Verbs (4.2.2.4.3) such as sleeping and napping are also subgroups of these verbs as classified by Levin. All these verbs have commonly-used non-PCP frame equivalents as well, but there are semantic nuances that make the speaker choose one or the other. Like psych verbs that use a different light verb (or a simplex verb) in the neutral diathesis, verbs involving the body have equivalents like *âroq zadan* ‘to burp’ [burp hit] for *âroq gereftan* ‘to feel like burping’ [burp take], *axm kardan* ‘to frown’ [frown do] for *axm gereftan* ‘to feel like frowning’ [frown take], *čort zadan* ‘to nap’ [nap hit] for *čort gereftan* ‘to feel like napping’ [nap take], *xâridan* ‘to itch’ for *xâreš gereftan* ‘to feel like itching’ [itching take], *larz kardan* ‘to shudder’ [shudder do] for *larz gereftan* [shudder take], etc.

4.2.2.4.1 Verbs of Bodily Process

Class Members: *âroq gereftan* ‘to feel like burping’ [burp take], *sekseke gereftan* ‘to get hiccups’ [hiccup take], *atse gereftan/umadan* ‘to get sneezy’ [sneeze take/come], *xamiyâze gereftan* ‘to feel like yawning’ [yawn take], *estefrâq gereftan* ‘to feel like vomiting’ [vomit take], *oq gereftan/umadan* ‘to feel like puking’ [puke take], *sorfe gereftan* ‘to get cough’ [cough take]

4.2.2.4.2 Verbs of Nonverbal Expression

Class Members: *axm gereftan* ‘to feel like frowning’ [frown take], *qor umadan* ‘to feel like growling’ [growl come], *xande gereftan/umadan* ‘to feel like laughing’ [laughter take/come], *gerye gereftan/umadan* ‘to feel like weeping’ [weeping take/come], *heqheq gereftan* ‘to sob’ [sobbing take]

4.2.2.4.3 Snooze Verbs

Class Members: *xâb umadan/bordan/gereftan* ‘to feel like sleeping’ [sleep come/carry/take], *čort gereftan/bordan* ‘to feel like napping’ [nap take/carry]

4.2.2.4.4 Verbs of Bodily State and Damage to the Body

4.2.2.4.4.1 Pain Verbs

Class Members: *dard umadan/gereftan* ‘to feel pain’ [pain come/take], *xâreš gereftan* ‘to feel itchy’ [itching take], *delpiče gereftan* ‘to get cramps’ [cramps take]

4.2.2.4.4.2 Verbs of (Change of) Bodily State

Class Members: *larz gereftan* ‘to shudder’ [shudder take], *za’f gereftan* ‘to faint’ [faint take], *tešne šodan/budan* ‘to get/be thirsty’ [thirsty become/be], *gorosne šodan/budan* ‘to get/be hungry’ [hungry become/be], *sard šodan/budan* ‘to get to feel/feel cold’ [cold become/be], *garm šodan/budan* ‘to get to feel/feel hot’ [hot become/be]

4.2.2.4.5 Dual Physical/Psych Predicates

The main denotation of these verbs is the physical sense, but it also has a mental sense as well. For example, one might vomit either physically (due to sickness) or (metaphorically) mentally (due to experiencing a stimulus that one resents).

Class Members: *estefrâq gereftan* ‘to feel like vomiting’ [vomit take], *larz gereftan* ‘to shudder’ [shudder take], *dard gereftan/umadan* ‘to feel pain’ [pain take/come], *sard*

šodan ‘to feel cold’ [cold become], *oq gereftan/umadan* ‘to feel like puking’ [puke take/come], *čendeš šodan* ‘to feel goose-bumps’ [goose-bumps become], *qelqelak gereftan/umadan/šodan* ‘to get tickled’ [tickle take/come/become] , *xâreš gereftan* ‘to feel itchy’ [itching take]

4.2.2.5 Verbs of Disappearance

Class Members: *qeib zadan* ‘to disappear’ [invisible hit]

4.2.2.6 Correspond Verbs

Class Members: *tavâfoq šodan* ‘to agree’ [agreement become], *bahs šodan* ‘to argue’ [argument become], *mo’âmele šodan* ‘to bargain’ [bargain become], *jarrobahs šodan* ‘to bicker’ [bicker become], *jang šodan* ‘to combat’ [battle become], *harf šodan/gereftan* ‘to confabulate’ [utterance become/take], *da’vâ šodan* ‘to quarrel’ [quarrel become]

Correspond verbs actually denote reciprocal events. They are the result of a diathesis alternation in some complex predicates that are reciprocal by nature (i.e. more than one person is needed for the action to happen), but one of the involved parties has more volition and intention compared to other parties, which means it has a higher degree of agentivity. These inherently reciprocal events use the light verb *kardan* ‘to do’ to highlight the agentivity of the person(s) with more volition, and the other involved parties with less volition are realized either in a comitative oblique form (a PP with the preposition *bâ* ‘with’, e.g. sentence 4-15), or in an accusative *râ*-marked form.

The diathesis alternation happens when an additional [-volition] semantic layer is intended to be added to the main involved participant. As a result, the PCP frame is used and the light verbs denoting agentivity will be substituted by light verbs denoting unintentionality of the participants and accidentalness of the event (*kardan* ‘to do’ → *šodan* ‘to happen’*; zadan* ‘to say’ → *gereftan* ‘to get engaged’).

da'vâ=m šod-Ø ba Ali
quarrel=1SG happen.PST.3SG.AGR with Ali
Quarrel happened between me and Ali.

(4-15)

There is also a situation where all the parties are equally involved. If the reciprocal event has happened intentionally ([+volition]), the regular syntactic frame is used, with the light verb agreeing with the morphosyntactic properties of the involved parties (one of the plural subject agreement markers). If the reciprocal event has happened unintentionally ([-volition]), the PCP frame is used with one of the light verbs denoting unintentionality of the participants.

4.2.2.7 Verbs of Relation

By definition, verbs of relation are verbs that imply that the referent of the compound subject of the sentence is larger or smaller than, dominant over, subordinate to, causative of, or resultant from some other referent (Erikson 1986).

In standard Persian, predicates like 'X is small for you' are expressed, comparably to English, by means of the preposition *barâye* 'for'. The glossed examples (4-16) and (4-17) respectively show the full form and clitic form pronoun variants.

in buluz barâye to bozorg ast-Ø
this shirt for 2SG.FPR big be.PRS-3SG.AGR
This shirt is big for you.

(4-16)

in buluz barây=at bozorg ast-Ø
this shirt for=2SG big be.PRS-3SG.AGR
This shirt is big for you.

(4-17)

Converted to PCP frame in colloquial Persian, this will become:

in buluz bozorg=et=e

this shirt big=2SG=be.PRS.3SG.AGR

This shirt is big for you.

(4-18)

Like all PCP constructions, the pronominal clitic represents the centered self which could be best described as having “dative” case.

4.3 Light-Verb-Based Classification

Despite the view that recognizes light verbs as semantically empty elements that function as a means to express the tense-aspect-mood features of the clause, light verbs in PCP construction play a much more significant role and semantically contribute to the denotation of the whole clause. In most cases, a less-frequent less-known denotation of the verb is utilized to compose the complex predicate (Samvelian 2012). For that matter, it is highly difficult to tell whether the verb used in a complex predicate is actually a light verb or a lexical verb. Despite this difficulty, I tried to find the less-known denotations of the verbs used in PCP construction. The following is the findings and results of this investigation.

The light verbs used to make compound verbs in any language are limited, and Persian is not an exception. The construction studied in this work (PCP) uses a subset of the set of light verbs in Persian. Table 4-2 is the list of (infinitive forms of) light verbs used in PCPs.

Table 4-2: the most frequent light verbs used in PCPs

LV		LV		LV		LV	
umadan	‘to come’	gereftan	‘to take’	bordan	‘to carry’	zadan	‘to hit’
šodan	‘to become’	budan	‘to be’				

There are a few light verbs that I want to include in a separate table (Table 4-3) since they are observed much less frequently (either in PCP-type verbs of memory or in verbs of desire).

Table 4-3: less frequent light verbs used in some limited PCPs

LV		LV		LV		LV	
raftan	'to go'	oftâdan	'to fall'	mundan	'to stay'	kešidan	'to pull/drag'

According to Dehkhoda Persian monolingual dictionary, the number of complex predicates[5] generated with these light verbs is:

LV		Frequency	LV		Frequency
umadan	'to come'	275	gereftan	'to take'	459
bordan	'to carry'	166	zadan	'to hit'	856
šodan	'to become'	500	budan	'to be'	159

Zadan 'to hit' leads the set of light verbs in being the most productive complex predicate generator. However, this is not the case with PCPs. Although there is no way to look up PCPs in Dehkhoda dictionary, a review of the past literature added to my own native speaker knowledge of Persian (supported by data from the Web) tells that *zadan* is among the least used light verbs used in the PCP construction. On the other hand, *gereftan* 'to take', and to a lesser extent *umadan* 'to come', are on top of the productivity list. Of the two copulas, *šodan* is the more productive one.

In this section, I will try to extract the finer semantic details of PCPs, trying to find a pattern that will reveal how light verbs function in PCPs. For this purpose, I will only depend on the data retrieved from the Web as my corpus. For the purpose of reference, I

[5] The preverbal element can be a noun phrase, an adjective phrase, a preposition, a prepositional phrase, or an adverb.

will include the conjugation of each verb in past and present tenses in the initial part of each subsection[6].

4.3.1 The Copulas: *šodan/budan*

	past		present		irrealis	
šodan	SG	PL	SG	PL	SG	PL
1	šodam	šodim	mišam	mišim	bešam	bešim
2	šodi	šodin	miši	mišin	beši	bešin
3	šod	šodan	miše	mišan	beše	bešan

	past		present[7]				irrealis	
budan	SG	PL	SG		PL		SG	PL
1	budam	budim	hastam	am	hastim	im	bâšam	bâšim
2	budi	budin	hasti	i	hastin	in	bâši	bâšin
3	bud	budan	hast	e	hastan	an	bâše	bâšan

Šodan is a change of state copula and a non-copulative light verb as well. In PCPs, it functions as a change of state copula in verbs of relation and possessive dative constructions, and as a light verb with two different denotations found in Dehkhoda dictionary.

---

[6] The 3rd person plural past of each verb conjugation might seem identical with the infinitive form, but actually in the infinitive form the stress is on the final syllable and in the past tense, the stress is on the initial syllable in case the verb is highlighted, and stressless in a complex predicate where the stress is on the preverbal element.

[7] The present tense conjugation of the copula ‘to be’ has a full form and a clitic form in colloquial Persian. I have shown the full form in the left column and the clitic form in the right one.

- To emerge (feeling of X emerged in the experiencer) [in psych verbs and a few change of bodily state verbs]
- To happen, to occur (X happened between us) [in correspond verbs]

According to the corpus data, the 'emerging' denotation of *šodan* is mostly used in a psych predicate, but it is also observed in a few verbs of change of bodily state. In a psych predicate, this state is a noun (such as gloss 4-19), and in verbs of change of bodily state, it is an adjective (such as gloss 4-20). In case of a noun preverbal element, the predicate would mean "The feeling of N is emerging in the experiencer", and in case of an adjective preverbal element, the predicate would mean "the feeling of [N derived from Adj meaning 'being Adj'] is emerging in the experiencer". In both cases, the experiencer is, by implication, a locus in which the feeling is emerging.

šarm=am šod-Ø (4-19)
abashment=1SG become.PST-3SG.AGR
I felt abashed.

gorosna=m šod-Ø
hungry=1SG become.PST-3SG.AGR
I felt hungry. (4-20)

Like other PCPs, these predicates have non-PCP neutral-diathesis alternatives. However, these are not exactly interchangeable with no change in meaning. Semantically, the affectedness of the experiencer is highlighted in the PCP alternative. With respect to argument structure and argument linking, the stimulus is more frequently realized as an

oblique argument (PP or CP_{inf}) in the PCP version, and more often with an adverbial clause in the non-PCP version[8].

About 500 complex predicates in Persian are composed using the light verb *kardan* 'to do' which indicates agentivity. Some of these verbs are psych verbs, such as *hasudi kardan* 'to envy' or *qaribi kardan* 'to feel alienated'. With these verbs, diathesis alternation can occur for the purpose of highlighting the affectedness and volitionlessness of the experiencer. The experiencer is being deprived of volition and s/he is being struck by the feeling, making *hasudi=m šod* 'the feeling of envy emerged in me' and *qaribi=m šod* 'the feeling of alienation emerged in me'.

There is yet another function for šodan, and that is in decausativization (see the diathesis alternation section in Chapter 3). In some transitive complex predicates, the agent imposes a state on the experiencer (which is realized as an oblique argument). In other words, the agent does something (which is not necessarily known) that causes the patient to get affected by it and experience a change of state. For the purpose of changing the diathesis and defocusing the agent, a decausativization operation is used to demote the agent argument and promote the experiencer to the main and only complement of the predicate, resulting in an intransitive predicate which indicates a change of state occurring to the experiencer. The light verb in the transitive predicate is usually *kardan* 'to do/to make' which denotes agentivity, but it changes to *šodan* in the decausativized intransitive predicate.

safar-ro	behe=m	kuft	kard-an	→	safar	kuft=am	šod-Ø.
trip-DOM	to=1SG.OBL	poison	make.PST-3PL.AGR		trip	poison=1SG.DAT	become.pst-3SG.AGR

They made the trip bitter for me. → the trip was bittered for me.

(4-21)

[8] This is based on corpus data

Budan ‘to be’ is a stative copula whose function, in PCPs, could be to express the steady presence of a psychological or physical state in the experiencer. The other usage it has is in verbs of relation to express an evaluative relation between two arguments, comparing the size of the subject argument with the capacity of a centered self which is the dative object argument. The centered self can either be realized in an oblique argument (a prepositional phrase using the preposition ‘for’) or in a PCP frame, as a pronominal clitic suffixed to the adjective describing the relation between the two arguments. The subject could be either a DP or a CP_{inf}.

In general, any complex predicate with the copula *budan* can rightfully use the copula *šodan* to indicate the changing nature of the state, implying that the state has been absent in the experiencer before and is now present in him/her. This is not surprising since, as mentioned above, one of the denotations of the verb *šodan* is ‘to change, to change from one state to another’. This general usage accounts for the function of *šodan* in (change of) bodily states, possessive dative constructions, and verbs of relation, all of which are eligible to use both *budan* and *šodan*.

4.3.2 Other Light Verbs

4.3.2.1 gereftan (‘to take’)

	past		present		irrealis	
gereftan	SG	PL	SG	PL	SG	PL
1	gereftam	gereftim	migiram	migirim	begiram	begirim
2	gerefti	gereftin	migiri	migirin	begiri	begirin
3	gereft	gereftan	migire	migiran	begire	begiran

According to Merriam-Webster online dictionary, the verb ‘take’ has 20 transitive denotations (with each denotation including a number of finer sub-denotations), and 8 different intransitive denotations. The Persian equivalent of ‘to take’ is *gereftan* which

has almost the same denotation diversity, and even more so in classical and literary Persian. This verb also takes a number of prepositions and makes complex predicates, each of which has a number of denotations per se. The light verb *gereftan* used in PCP constructions is actually derived from the classical Persian *dar-gereftan* [in-take] and *bâz-gereftan* [again-take], but the preposition is never used[9]. Looking up the denotations of the compound verbs that use *gereftan* in Dehkhoda dictionary, I found the following three denotations that can account for the functions *gereftan* has as a light verb used in PCPs. These three denotations include:

- *dar-gereftan*: to occupy, to seize, to capture [used in psych verbs and verbs involving body]
- *bâz-gereftan*: to impel, to urge [used productively, taking a simple or phrasal gerundive complement, denoting "feeling an urge by the experiencer to do the action denoted by the gerund"]
- *dar-gereftan*: (with nouns denoting talking) to engage in an agreeable loose conversation.

Two example sentences for each nuance/function are glossed below.

Psych verbs:

kasi	az	na-didan-e	to	qossa=š	ne-mi-gir-e.
anyone	from	NEG-seeing-EZF	2SG.FPR	sorrow=3SG	NEG-IND-occupy.PRS-3SG.AGR

Nobody gets sad by not seeing you. (Lit. nobody gets occupied by sorrow by not seeing you.)

(4-22)

[9] Native speakers of Modern Persian are unaware of this fact about this light verb if they don't delve into Persian literary classics or refer to dictionaries.

bâz	ham	ke	az	bolandi	tars=et	gereft-Ø!
again	again	that	from	height	fear=2SG	occupy.PST-3SG.AGR

Again you got afraid of height. (Lit. again you got occupied by fear from height.)

(4-23)

Verbs involving body:

moqe'-e	dars	xundan	xâb=am	mi-gir-e.	
time-EZF	lesson	studying	sleep=1SG	IND-occupy.PRS-3SG.AGR	

I get sleepy when studying. (Lit. I get seized by sleep when studying.)

(4-24)

man	age	az	tah-e	del	be-xand-am	sekseka=m	mi-gir-e.
1SG.FPR	if	from	bottom-EZF	heart	SBJV-laugh.PST-1SG.AGR	hiccups=1SG	IND-occupy.PRS-3SG.AGR

I would get hiccups if I laugh from the bottom of my heart. (Lit. I get seized by hiccups if I laugh from the bottom of my heart.)

(4-25)

Productive usage:

šab-e	âxar-i	masxare-bâzi=m	gerefte.
night-EZF	last-DEF	goofiness=1SG	urge.PRS.PRF

Now in the last night, I feel an urge for being goofy. (Lit. I'm urged to do goofiness.)

(4-26)

Hâlâ	čerâ	qesse	goftan=am	gerefte?
now	why	story	telling=1SG	urge.PRS.PRF

(I'm wondering) why I'm feeling an urge for story-telling? (why I'm urged to do story-telling?)

(4-27)

Talking:

sohbat=emun	gereft-Ø		o	qadam	zad-im.
talk=1PL	engage.PST-3SG.AGR		and	stride	move.PST-1PL.AGR

We got engaged in an agreeable conversation and strolled.

(4-28)

vaqti	mi-r-am		yeho	harf=etun	mi-gir-e
when	IND-go.PRS-1SG.AGR		suddenly	talk=2PL	IND-engage.PRS-3SG.AGR

When I go, y'all suddenly get engaged in an agreeable conversation.

(4-29)

There are 37 PCP constructions in our corpus that use the light verb *gereftan*, all of which I have listed in the class member part of section 4.2, and they are also listed in Appendix-1 and Appendix-2.

4.3.2.2 umadan ('to come')

	past		present		irrealis	
umadan	SG	PL	SG	PL	SG	PL
1	umadam	umadim	miâm	miâim	biâm	biâim
2	umadi	umadin	miâi	miâin	biâi	biâin
3	umad	umadan	miâd	miân	biâd	biân

umadan 'to come' has a large and diverse list of denotations and usages in producing complex predicates in Persian. According to Dehkhoda dictionary, one of the main denotations of *umadan* which is widely used in producing complex predicates in Classical Persian is 'to feel'. An example from Shahnameh (Ferdowsi, c. 977-1010 CE) is provided in (4-30).

hami	az	šomâ	in	šegeft	ây-ad=am
Indeed	from	2PL.FPR	this	amazement	come.PRS-3SG.AGR=1SG.DAT

Indeed I feel amazed that you did this. (Lit. Indeed amazement is coming to me by this from you.)

(4-30)

This light verb is used in psych predicates and verbs of bodily process using the PCP construction with the denotation ‘to feel’. However, it should be noted that there is a slight difference between the similar uses of *šodan* and *umadan*, like *šarmam šod* and *šarmam umad*. Although they both denote ‘I’m feeling ashamed’, in the *umadan* version there is an added sense of the experiencer being the target or recipient of the feeling. This is because *umadan* is a motion verb, and the sentence could have a metaphoric reading of “the feeling is coming toward the experiencer”.

Below are two glossed example sentences from our corpus.

xejâlat=am	umad-Ø	be-r-am	pâk=eš	kon-am.
embarrassment=1SG.DAT	come.PST-3SG.AGR	SBJV-go.PRS-1SG.AGR	delete=3SG.ACC	make.PRS-1SG.AGR

I felt embarrassed to go and delete it.

(4-31)

man	ke	aslan	xâb=am	ne-mi-â-d.
1SG.FPR	that	not at all	sleep=1SG.DAT	NEG-IND-come.PRS-3SG.AGR

As for me, I’m not feeling sleepy at all.

(4-32)

Umadan is actually another productive light verb like *gereftan*. In addition to psych verbs and verbs of bodily process, it can take any complement as long as the complement is a gerund denoting an action. Usually, if the gerund is derived from a compound verb, the preverbal nominal element is extracted and *umadan* is added to it to denote that the person feels in the mood for doing that act.

hâlâ	ke	konkur	dâd-am	hey	dars=am	mi-â-d.
now	that	exam	give.PST-1SG.AGR	all the time	lesson=1SG.DAT	IND-come.PRS-3SG.AGR

Now that I have given the exam, I feel in the mood for studying all the time.

(4-33)

In this example, *dars* ‘lesson’ is the nominal element used in the complex predicate *dars xundan* ‘to study’ [lesson read]. The glossed example (4-33) shows that this nominal element is extracted and used in isolation, while it retains the meaning of the whole complex predicate before getting extracted.

The psych verbs and verbs of bodily process that use the light verb *umadan* in a PCP frame are listed in Appendix-1.

4.3.2.3 bordan (‘to carry’)

	past		present		irrealis	
bordan	SG	PL	SG	PL	SG	PL
1	bordam	bordim	mibaram	mibarim	bebaram	bebarim
2	bordi	bordin	mibari	mibarin	bebari	bebarin
3	bord	bordan	mibare	mibaran	bebare	bebaran

The light verb *bordan* is an unproductive light verb in PCP constructions. It is used only with conjecture verbs, snooze verbs, and marvel verbs (as far as found in the corpus). There is also a very limited (probably stylistic) use of *bordan* in a subset of admire verbs which I call **resentment verbs**. These three verbs include: *zur bordan* ‘to begrudge’, *laj bordan* ‘to feel pissed off’, and *hers bordan* ‘to feel peeved’. The frequency of their occurrence is so low in the corpus.

According to Dehkhoda dictionary, the first denotation of *bordan* is ‘to carry, to move’, the antonym being ‘to bring’. The other denotation is *šodan* ‘to become’ which is

the change of state copula. The first denotation is used in conjecture verbs and the second one is used in snooze verbs and marvel verbs.

In conjecture verbs the experiencer is subject and it has genitive case. The person is somehow the initiator or the source of the action, and his/her conjecture which is a mental capacity owned by him/her moves like an arrow to be landed on the target about which s/he is suspicious. So the main denotation of *bordan* is metaphorically used in these verbs, with a cognitive motion intended.

In snooze verbs and marvel verbs, the experiencer is a dative object since s/he is affected / struck by the feeling, and not the initiator of any action. The second denotation of *bordan* is used for these verbs, expressing a change of state in the experiencer, but it is slightly different from using the change of state copula in that it is metaphorically taking from the heavy verb *bordan* 'to carry, to move', connoting "the experiencer being carried/moved, subconsciously leaving his/her current mental situation, as a result of getting affected by the experience". Below I have provided an example for each of the verb classes that use this light verb.

Snooze Verb

har-do=mun	az	xastegi	xâb=emun	bord-Ø.
any-two=1PL	because of	fatigue	sleep=1PL.DAT	carry.PST-3SG.AGR

Both of us fell asleep because of fatigue.

(4-34)

Marvel Verb

vaqti	man-o	did-Ø	yeho	mât=eš	bord-Ø.
when	1SG.FPR-DOM	see.PST-3SG.AGR	suddenly	aghast=3SG.DAT	carry.PST-3SG.AGR

When s/he saw me, s/he suddenly got aghast.

(4-35)

Conjecture Verb

dâr-e kam-kam be xod=am šakk=am mi-bar-e.
PROG-3SG.AGR little-little to self=1SG suspicion=1SG.GEN IND-carry.PRS-3SG.AGR
I'm growing suspicious of myself. (Lit. my suspicion is gradually carried to myself.)
(4-36)

4.3.2.4 zadan ('to hit, to strike')

	past		present		irrealis	
zadan	SG	PL	SG	PL	SG	PL
1	zadam	zadim	mizanam	mizanim	bezanam	bezanim
2	zadi	zadin	mizani	mizanin	bezani	bezanin
3	zad	zadan	mizane	mizanan	bezane	bezanan

Zadan is the most widely used light verb in making complex predicates in Persian. As mentioned earlier, there are about 850 complex predicates and expressions using this light verb. PCP construction, however, does not use this light verb productively as in non-PCP complex predicates. *zadan* might have various denotations, and it's not necessarily its main denotation, 'to hit, to strike', that is used to generate PCPs.

This light verb is observed in marvel verbs and verbs of disappearance (as far as found in the corpus). In marvel verbs, it implies a harsher affectedness of the experiencer, as if s/he has been struck or even petrified by the experience caused by the stimulus.

The other point that should be remarked is that the light verb *zadan* could also be considered as a lexical aspect, indicating the suddenness of the action or change of state. This is relevant both to marvel verbs and (particularly so to) verbs of disappearance, all of which imply suddenness and unexpectedness of what is happening.

az	šenidan-e	sedâ=š	xošk=eš	zad-Ø.
from	hearing-EZF	voice=3SG.GEN	stiff=3SG.DAT	strike.PST-3SG.AGR

S/he (suddenly) got astonished by hearing his/her voice. (Lit. s/he got struck by astonishment…)

(4-37)

4.4 Summary

In this chapter I provided a rather detailed analysis of the PCP construction data found in the corpus. Drawing on Levin's (1993) insights, I looked at the argument structure of each verb, compared it with allosentences which are realized in the regular non-PCP syntactic frame. This helped me discover the semantic nuances encoded in the PCP construction.

Based on this, I presented a lexical semantic classification of the data in section 4.2. The verbs were classified into those which take a possessor centered self as their main argument (section 4.2.1) and those which take a dative centered self (section 4.2.2).

In section 4.3, I classified PCPs based on the light verbs they use in their composition. It was noted that the light verb significantly contributes to the semantic content of the predicate, and I found denotations of the verbs that show this contribution.

CHAPTER 5 - APPLICATION OF THEORY TO PERSIAN DATA

5.1 Diathesis Alternation

According to the definitions provided in Chapter 3, the construction studied in this work could rightfully be labeled either as a periphrastic construction or a complex predicate. Bonami and Samvelian (2009) analyze five types of periphrastic constructions in Persian. But the PCP construction is not among them.

The PCP construction is a particular morphosyntactic form which, in line with the notion of diathesis, serves a particular purpose in expressing the speaker's construal of an event or state. These constructions are clearly anthropocentric, classifying the whole world into (a) the human who is at the center, and (b) everything else. In other words, the speaker's construal of the situation is that "self" is the center, and the "external world", i.e. everything other than the centered self (human), has a relation to this center. It's a matter of being in or out with respect to the subconscious circle of a person. And in this self-centered view, one might sometimes want to express oneself as having a non-agentive status with regards to the external world. For example, you might get affected by an external (concrete or abstract) stimulus, or look at an emotion or state as coming on you of "its" own volition, as it were (Croft 1993: 66). Or an external object or state might have a certain relation with you that you have no control on, e.g. 'X is <u>yours</u>' (dative possessive constructions) or 'as for <u>me</u>, X is Y' where X is an external object and Y is the relation of X compared to 'me' as the centered self (verbs of relation). Such a view makes the human (which is the chief nominal participant in the situation) be construed as

a target or locus affected by stimuli from the external world. In case of stative mental verbs where there is no transmission of force (Croft 1993), the centered self which is the experiencer is still getting affected by an external stimulus, even though the stimulus is not expressed or implied. So the PCP construction is used to convey the affectedness of the centered self.

In a self-centered perspective, the centered self could be interpreted as being volitionless. So, in an inherently intransitive event like sleeping, if the speaker wants to highlight the unintentionality of the event (i.e. to say that "the feeling of sleepiness happened to me", or "I fell asleep" vs. "I decidedly went to sleep"), a PCP construction (volitionless diathesis) (example 5-2) is used instead of a regular predicate (neutral diathesis) (example 5-1).

yek sa'at piš xâbid-am
one hour ago sleep.PST-1SG.AGR
I slept an hour ago.

(5-1)

yek sa'at piš xâb=am gereft-Ø
one hour ago sleep=1SG.DAT occupy.PST-3SG.AGR
I got sleepy an hour ago (Lit: Sleeping occupied me an hour ago.)

(5-2)

By the same token, showing unintentionality of the centered self in a reciprocal event (example 5-3) would be achieved by means of volitionless diathesis (example 5-4).

diruz da'vâ kard-im
yesterday quarrel do.PST-1PL.AGR
We quarreled yesterday.

(5-3)

diruz da’vâ=mun šod-Ø

yesterday quarrel=1PL.DAT become.PST-3SG.AGR

We (unintentionally) quarreled yesterday. (Lit. Quarrel happened to us yesterday)

(5-4)

Like middle which is a diathesis cluster, PCP construction represents a diathesis cluster as well, including diatheses that intend to convey (a) the affectedness of an experiencer (centered self), (b) the volitionlessness of the centered self, (c) some external (concrete or abstract) object evaluated (for some feature) against the (capacity of) centered self. In all these cases, the centered self is non-agentive, and also not highly affected like a patient. In this regard, it is to some extent similar to the middle diathesis, but they are not exactly the same.

Verbs of perception (including verbs of memory, verbs of desire, and conjecture verbs) are the only PCPs in which the pronominal clitic is analyzed as having genitive case. In these verbs, a part of human perceptual power is taken by the centered self as being so close to his/her subconscious, being in an inalienable possessive relation with it. An object in the external world is perceived in a particular manner, not by the centered self, but by a part of his/her mind inalienably possessed by him/her. So, although there seems to be some degree of volition in the centered self, it is not a direct participant of the clause.

The PCP construction is an instance of External Possessor Construction proposed by Haig (2008). The main participant, which is the most eligible argument to be the topic of the clause, is an indirect participant (cf. Figure 3-1). Indirect participation could be attributed to several semantic roles such as experiencer, benefactive, and possessor. They do not have control over the event or situation, but the event or situation is highly relevant to them. The bundle of semantic roles that are considered indirect participation is the functional domain of the Old Persian Genitive, which comprises of dative and genitive cases.

Based on the analysis provided in Chapter 4, the main participant in a PCP construction is an indirect participant and it is represented by genitive case, which is a unification of dative and genitive in Persian. The main participant (that gets topic-shifted) is possessor in possessor centered-self category, experiencer in psych predicates, long verbs, verbs involving the body, and correspond verbs, and benefactive in verbs of relation. In all of them, the main participant is highly relevant to the event or situation although it is not directly involved, i.e. it does not have control over the event or situation as an agent does, and it is not as affected as a patient. The Levin-based verb classification in Chapter 3 clearly divides the PCP construction into dative and possessive (i.e. genitive) centered self. The centered self, which is the main participant of the clause in PCP construction is an indirect participant, which according to Haig (2008) is represented by the functional domain of genitive (which is the single case expressing both dative and genitive in indirect participation). As Haig suggests they are all sentient beings, most commonly human beings, and this is what makes them highly eligible to be the topic of the clause.

5.2 Agreement

The previous works on PCP construction either say that there is no subject-verb agreement or that the pronominal clitic is a subject-agreement marker (Rasekhmahand 2010), or that the subject is not the experiencer but the theme argument which is always 3rd person singular and the verb always agrees with it (Dabir-Moghaddam 1997, Sedighi 2005). I propose a different analysis which is in part in line with the third view.

These complex predicates are composed of a theme as the preverbal element, and a light verb. The theme is a state or an event, not a human being. So it is inherently 3rd person singular. The diathesis adopted in the PCP construction makes this theme nominal element the subject of the sentence (probably to show the indirectness of the participation

of the main participant). The verb always agrees with this subject, hence it is always in 3rd person singular conjugation form.

I adopt Kim's (2004) idea to propose a hybrid dual-layer agreement including a grammatical or morphosyntactic layer of agreement, and a semantic layer of agreement. The theme is the subject and triggers grammatical agreement on the verb. The semantic layer of agreement is with the indirect participant. In the PCP construction, the indirect participant is a sentient being, most commonly [+human], which is highly likely to be the topic of the clause according to Givón's (1976) topicality hierarchy. If Givón's idea on the true nature of agreement is accepted, it's only natural that the indirect participant would trigger agreement in the clause. Since the indirect participant is in the functional domain of genitive in Persian, the paradigm of oblique pronominal enclitic is being used to mark it. The pronominal enclitics, which are phrasal affixes, have been full form pronouns in Classical Persian. This is in line with Givón's second argument that agreement markers have been pronouns which have been reanalyzed. This reanalysis is observed in PCP construction over time, grammaticalizing the pronouns to pronominal enclitics which are actually edge inflection suffixes. The position they are realized in the clause is in fact where object clitics in other (non-PCP) functions appear, i.e. in the second position.

The optionality of the clause-initial DP, i.e. the indirect participant which is the topic, is simply because Persian is a pro-drop language. In fact, this is another reason to accept the argument that the pronominal enclitics are actually agreement markers, because it's only in the presence of an agreement system that a language could be pro-drop, where the topic is dropped when the speaker is certain that dropping would not lead to any information loss or miscommunication. The likelihood of the topic being dropped in such a system is almost equal to the likelihood of an argument being eligible as topic based on Givón's topicality hierarchy. 1st person is most likely to be topic and pro-dropped because of the ego-centric nature of discourse. 2nd person has the next high likelihood because the addressee is present in the discourse and certainly knows whom the speaker is addressing. The least likelihood of an argument being topic and dropped belongs to 3rd

person. In fact it gets dropped only in cases where the speaker is certain that the absent argument is commonly present in the consciousness of both sides of the dialogue.

The type of agreement in PCP construction, I'd say, is the same type of agreement in the neutral syntax of Persian which leads to having a pro-drop system. Tentatively, it is a registration-type agreement which does not violate Principle B of Binding Theory. The agreement operation includes only feature copying and no co-indexation.

5.3 Pronominal Clitics

Our analysis of pronominal clitics in the PCP construction as being agreement markers (in the particular diathesis cluster that triggers the use of a PCP frame) is in line with Anderson's (2005) proposal that they are actually overt morphological markers of the morphosyntactic properties of a phrase. They behave like inflectional morphological markers as part of the functional content of the clause.

This analysis is only applicable to colloquial Persian, and I argue that the pronominal clitics have not yet been completely reanalyzed as morphological phrasal affixes in literary Persian which still retains some features of Classical Persian. Such forms can be seen even in social networks such as Facebook today and it's perfectly comprehended, but only by literate readers who are familiar with Classical Persian literature. Sentence (5-5) is an example from Facebook. The author intends to give a literary aura to her writing, so she uses this construction which is not used in modern colloquial Persian to achieve her purpose, making it more literary and consequently more romantic. The non-literary contemporary colloquial equivalents follow. The centered self (1SG) is marked with differential object marker in the literary style, and in the modern colloquial forms it could either be realized as an oblique PP (5-6), or using a PCP frame (5-7).

budan-e to ma-râ bas ast-Ø

being-EZF 2SG.FPR 1SG.FPR-DOM enough be.PRS.3SG.AGR

It’s enough for me that you are present. (Lit. being of you is enough for me)

(5-5)

budan-e to barâye man bass=e

being-EZF 2SG.FPR for 1SG.FPR enough=be.PRS.3SG.AGR

It’s enough for me that you are present. (Lit. being of you is enough for me)

(5-6)

Or

budan-e to bass=am=e

being-EZF 2SG.FPR enough=1SG.DAT=be.PRS.3SG.AGR

It’s enough for me that you are present. (Lit. being of you is enough for me)

(5-7)

CHAPTER 6 – CONCLUSION

6.1 Summary and Conclusion

The construction introduced in this work is a special non-neutral syntactic pattern found in a number of situations. I call this construction *Pronominal Complex Predicate* because it is the only type of complex predicate in colloquial Persian where the presence of a pronominal clitic is obligatory. So this naming is not based on any semantic criteria. The definition I assumed for a complex predicate is not different from a periphrastic construction. So they could justly be called either a *complex predicate* or a *periphrastic construction*. Therefore, I can say that the analysis presented in this work could be a contribution to the diversity of Persian periphrastic constructions suggested by Bonami and Samvelian (2009).

I analyzed the PCP construction as a syntactic sequence which is used to represent a special indirect participation diathesis adopted by the speaker. In such a perspective, the speaker alienates the person whom the sentence is about (or in my words “the centered self”) from the external world, and puts the centered self in a non-agentive status with regard to the external world, taking a volitionless position. Such a worldview constructs a diathesis cluster which results in using the PCP frame for expressing the sentence.

All the issues raised in the introduction chapter were answered by recourse to theories developed by Kulikov (2011), Givón (1976), Andeson (2005) and Haig (2008) as groundwork. The pronominal clitic was analyzed as an agreement marker in PCP constructions, being an inflectional morphological property of the clause. The centered

self which is topicalized is realized in the sentence-initial position, and can be optionally dropped since there is an agreement marker that registers it in the sentence.

A diachronic study is certainly necessary and gives valuable clues for a synchronic study. But it's definitely not adequate for an all-encompassing analysis. Using the web as a corpus paced up the data collection process and enabled us to find a more diverse array of data. This work is the first study that presents a unified analysis for such a wide range of data.

6.2 Recommendations for further study

During the course of this study, a number of questions occurred to me that require further investigation. I have listed the topics of these questions below.

- Compositionality of Persian complex predicates in general, and PCP construction in particular
- A cognitive first language acquisition study on how complex predicates are learned, and on what basis they are coined and why the coined ones are readily comprehended by native speakers who have never heard them before.
- How can such a complex idiosyncratic construction be taught to second language learners of Persian
- How can this theory get implemented in a computer program such that the number of ungrammatical sentences the computer program might generate is minimized

Appendix-1 The PCPs in the corpus: light-verb-based classification

gereftan	**umadan**	**šodan**	**budan**	**zadan**	**bordan**	**kešidan**
zur	yâd	edde’â	yâd	xošk	šakk	meyl
âr	xoš	čand sâl	čand sâl	boht	zann	ešq
estefrâq	bad	âr	qosse	mât	gamun	
hers	zur	šarm	tešne	qeib	hads	
laj	âr	xejâlat	gorosne		xošk	
oq	hers	qosse	sard		boht	
xašm	laj	hasudi	garm		mât	
qeiz	oq	tešne	bass		xâb	
vahšat	ta’ajjob	gorosne	ziâdi		čort	
delšure	tars	sard	zahmat			
ta’ajjob	šarm	garm	edde’â			
tars	xejâlat	qelqelak				
šarm	dard	čendeš				
xejâlat	rahm	tavâfoq				
qorur	heif	bahs				
qosse	havas	mo’âmele				
larz	qor	jarrobahs				
dard	xande	jang				
boht	gerye	harf				
xâreš	xâb	da’vâ				
âroq	qelqelak	bass				
sekseke		ziâdi				
atse		kasre ša’n				
xamiyâze		zahmat				
sorfe		xaste				
axm		kuft				

xande gerye heqheq xâb čort delpiče za’f qelqelak harf šuxi da’vâ						

Appendix-2 Alphabetical ordering of the PCPs with glosses

preverbal element	gloss	light verb(s)	predicate gloss
âr	‘abashment’	gereftan, umadan, šodan	to feel abashed
âroq	‘burp’	gereftan	to feel like burping
atse	‘sneeze’	gereftan	to get sneezy
axm	‘frown’	gereftan	to feel like frowning
bad	'bad'	umadan	to dislike
bahs	'argument'	šodan	to argue
bass	'enough'	šodan, budan	to become/be enough for someone
boht	‘bewilderment’	gereftan, zadan, bordan	to get bewildered
čand sâl	'how many years'	šodan, budan	how old is someone
čendeš	'goose-bumps'	šodan	to feel goose-bumps
čort	‘nap’	gereftan, bordan	to feel like napping, to fall asleep
da’vâ	‘quarrel’	gereftan, šodan	to quarrel
dard	‘pain’	gereftan, umadan	to feel pain
delpiče	‘cramps’	gereftan	to get cramps
delšure	‘anxiety’	gereftan	to get anxious
edde’â	'claim'	šodan, budan	to claim
ešq	'love'	kešidan	to will (to do something); to want (to do something) at will
estefrâq	‘vomit’	gereftan	to feel like vomiting
gamun	'guess'	bordan	to guess
gerye	‘weeping’	gereftan, umadan	to feel like weeping
gorosne	'hungry'	šodan, budan	to get/be hungry
garm	'hot'	šodan, budan	to get/be feeling hot
hads	'conjecture'	bordan	to conjecture
harf	‘utterance’	gereftan, šodan	to confabulate
hasudi	'envy'	šodan	to envy

havas	'crave'	umadan	to crave
heif	'alack'	umadan	to regret
heqheq	‘sobbing’	gereftan	to sob
hers	‘peeve’	gereftan, umadan	to get peeved
jang	'battle'	šodan	to combat
jarrobahs	'bicker'	šodan	to bicker
kasre ša’n	'dishonor'	šodan	to get dishonored
kuft	'poison'	šodan	to be made bitter
laj	‘irritation’	gereftan, umadan	to feel pissed-off
larz	‘shudder’	gereftan	to shudder
mât	'astonished'	zadan, bordan	to get astonished
meyl	'desire'	kešidan	to desire (to do something)/crave (something)
mo’âmele	'bargain'	šodan	to bargain
oq	‘puke’	gereftan, umadan	to feel like puking
qeib	'invisible'	zadan	to disappear
qeiz	‘rage’	gereftan	to get enraged
qelqelak	‘tickle’	gereftan, umadan, šodan	to get tickled
qor	'growl'	umadan	to feel like growling
qorur	‘pride’	gereftan	to get proud
qosse	‘sorrow’	gereftan, umadan, šodan	to get sad
rahm	'mercy'	umadan	to pity
šakk	'suspicion'	bordan	to get suspicious
sard	'cold'	šodan, budan	to get to feel/feel cold
šarm	‘shame’	gereftan, umadan, šodan	to get ashamed
sekseke	‘hiccup’	gereftan	to get hiccups
sorfe	‘cough’	gereftan	to get cough
šuxi	‘joke’	gereftan	to feel like joking
ta’ajjob	‘amazement’	gereftan, umadan	to get amazed

tars	‘fear’	gereftan, umadan	to get afraid
tavâfoq	'agreement'	šodan	to agree
tešne	'thirsty'	šodan, budan	to get/be thirsty
vahšat	‘fear’	gereftan	to get afraid
xâb	‘sleep’	gereftan, umadan, bordan	to feel like sleeping/to fall asleep
xamiyâze	‘yawn’	gereftan	to feel like yawning
xande	‘laughter’	gereftan, umadan	to feel like laughing
xâreš	‘itching’	gereftan	to feel itchy
xašm	‘anger’	gereftan	to get angry
xaste	'tired'	šodan	to get tired
xejâlat	‘embarassment’	gereftan, umadan, šodan	to get embarrassed
xoš	'pleasantness'	umadan	to like
xošk	'stiff'	zadan, bordan	to be petrified
yâd	'memory'	umadan, budan	to remember
yâd	'memory'	raftan	to forget
za’f	‘faint’	gereftan	to faint
zahmat	'trouble'	šodan, budan	to become/be a trouble for someone
zann	'suspicion'	bordan	to get suspicious
ziâdi	'too much'	šodan, budan	to become/be too much for someone
zur	‘force’	gereftan, umadan	to begrudge

References

Ackerman, Farrell, and Gert Webelhuth. *A theory of predicates*. Vol. 76. Stanford: CSLI Publications, 1998.

Adger, David, and Gillian Ramchand. "Psych nouns and predication."*PROCEEDINGS-NELS*. Vol. 36. No. 1. 2006.

Alsina, Alex, Joan Wanda Bresnan, and Peter Sells, eds. *Complex predicates*. Center for the Study of Language and Information, 1997.

Alsina, Alex, Miriam Butt, and Tracy Holloway King. "A theory of structure-sharing: Focusing on long-distance dependencies and parasitic gaps."*Proceedings of the LFG08 Conference*. 2008.

Anderson, Stephen R. *Aspects of the Theory of Clitics*. Oxford University Press, USA, 2005.

Arefi, Ma'edeh. *Overview of Impersonal Constructions in Persian Language*. Thesis. Allameh Tabatabaei University, 2011.

Barjasteh, Darab. Morphology, syntax and semantics of Persian compound verbs: A lexicalist approach. Diss. University of Illinois at Urbana-Champaign, 1983.

Boeckx, Cedric. "Quirky agreement." *Studia Linguistica* 54.3 (2000): 354-380.

Bonami, Olivier, and Pollet Samvelian. "Inflectional periphrasis in Persian."*Proceedings of the HPSG 2009 Conference*. 2009.

Bonami, Olivier, and Gert Webelhuth. "The phrase-structural diversity of periphrasis: a lexicalist account." *Periphrasis: The role of syntax and morphology in paradigms* (2013).

Croft, William. "Case marking and the semantics of mental verbs." *Semantics and the Lexicon*. Springer Netherlands, 1993. 55-72.

Dabir-Moghaddam, Mohammad. "Compound verbs in Persian." *Studies in the Linguistic Sciences* 27.2 (1997): 25-59.

Erikson, Kai. "The sociologist's hand." (1986): 808-811.

Ghomeshi, Jila. "Topics in Persian VPs." *Lingua* 102.2 (1997): 133-167.

Givón, Talmy. "Topic, pronoun and grammatical agreement." *Subject and topic*149 (1976): 188.

Givón, Talmy. *Syntax: an introduction*. Vol. 1. John Benjamins Publishing, 2001.

Goldberg, Adele E. "Words by default: Optimizing constraints and the Persian complex predicate." *Proceedings of the Annual meeting of the Berkeley Linguistics Society*. Vol. 22. No. 1. 2012.

Haig, Geoffrey LJ. Alignment change in Iranian languages: a construction grammar approach. Vol. 37. Walter de Gruyter, 2008.

Harley, Heidi Britton. *Subjects, events and licensing*. Diss. Massachusetts Institute of Technology, 1995.

Karimi, Simin. "Obliqueness, specificity, and discourse functions: Ra in Persian." *Linguistic Analysis* 20 (1990): 139-191.

Karimi, Simin. A Minimalist approach to scrambling: Evidence from Persian. Vol. 76. Walter de Gruyter, 2005.

Kemmer, Suzanne. *The middle voice*. Vol. 23. John Benjamins Publishing, 1993.

Kim, Jong-Bok. "Hybrid agreement in English." (2004): 1105-1128.

Kipper, Karin, Hoa Trang Dang, and Martha Palmer. "Class-based construction of a verb lexicon." *AAAI/IAAI*. 2000.

Khanlari, Parviz Natel. "Dastur-e Zaba'n-e Farsi [Persian Grammar]." *Tehran: Bonyad-e Farhang-e Iran* (1972).

Khanlari, Parviz. "Tarix-e zabân-e farsi (A History of the Persian Language)."*Tehran, Iran: Nashr-e No Publishing Company* (1986).

Khānlarī, Parvīz Nâtil. *Târîkh-i zabân-i Fârsî*. Vol. 1. Intishârât-i Bunyâd-i Farhang-i Irân, 1971.

Korn, Agnes. "Western Iranian Pronominal Clitics." *Orientalia Suecana* (2009): 159.

Kulikov, Leonid. "Voice typology." *The oxford handbook of linguistic typology*(2011): 368-398.

Lambrecht, Knud. Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Vol. 71. Cambridge university press, 1996.

Landau, Idan. *The locative syntax of experiencers*. Vol. 34. Cambridge, MA: MIT Press, 2010.

Lazard, Gilbert. *A grammar of contemporary Persian*. Costa Mesa, California: Mazda Publishers, 1992.

Levin, Beth. English verb classes and alternations: A preliminary investigation. University of Chicago press, 1993.

Lyons, John. *Introduction to theoretical linguistics*. Cambridge university press, 1968.

McCloskey, Jim. *Subjecthood and subject positions*. Springer Netherlands, 1997.

Pullum, Geoffrey K., and Arnold M. Zwicky. "Cliticization vs. inflection: English n't." *Language* 59.3 (1983): 502-513.

RASEKHMAHAND, M. "PERSIAN CLITICS BESIDE VERBS." *JOURNAL OF RESEARCHES IN LINGUISTICS* (2010).

Reinhart, Tanya. *Interface strategies*. Utrecht University, Research Institute for Language and Speech, 1995.

Samvelian, Pollet. Grammaire des prédicats complexes: les constructions nom-verbe. Lavoisier, 2012.

Samvelian, Pollet, and Jesse Tseng. "Persian object clitics and the syntax-morphology interface." Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar. 2010.

Sedighi, Anousha. *Quirky Subjects: Do They Exist in Persian?*. Diss. University of Ottawa, 2001.

Sedighi, Anousha. *Subject-predicate agreement restrictions in Persian*. Diss. University of Ottawa, 2005.

Sharifian, Farzad, and Ahmad R. Lotfi. "“When stones falls”: a conceptual–functional account of subject–verb agreement in Persian." *Language Sciences*29.6 (2007): 787-803.

Sigurðsson, Halldór Ármann. "The locus of case and agreement." *Working papers in Scandinavian syntax* 65 (2000): 65-108.

Talmy, Leonard. *Semantic structures in English and Atsugewi*. linguistics Department, University of California, 1972.

Van Belle, William, and Willy Van Langendonck, eds. *The Dative: Volume 1: Descriptive studies*. Vol. 2. John Benjamins Publishing, 1996.

Windfuhr, Gernot L. *Persian grammar: History and state of its study*. Vol. 12. Walter de Gruyter, 1979.

Zwicky, Arnold M. *On clitics*. Indiana University Linguistics Club, 1977.

تقی وحیدیان کامیار. *فعلهای یك شناسه*، نامه فرهنگستان زبان، دوره ٦، شماره ٢، سال ١٣٨٢

Vita

Ghazaleh Kazeminejad

Place of Birth

Tehran, Iran

Education

B.A. English Language and Literature, 2011
University of Tehran
Tehran, Iran

Professional Positions

2012-2014 Teaching Assistant, Linguistics Program
University of Kentucky
Lexington, Kentucky

2008-2011 Translator and Editor
University of Tehran Translation Center
Tehran, Iran

2007-2011 Translator and Editor
Adinehbook Publication Company
Tehran, Iran